DNA Hydroxymethylation and Methylation Dynamics during Neural Differentiation

Sabrina Karen Stewart

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

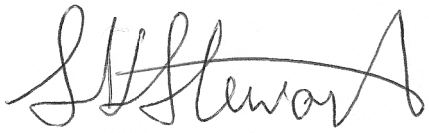
UCL Cancer Institute

University College London

2015

Declaration

I, Sabrina Karen Stewart, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



16th October 2015

Publications

1. **Stewart, S.K.**, Morris, T.J., Guilhamon, P., Bulstrode, H., Bachman, M., Balasubramanian, S., and Beck, S. (2015). oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. Methods *72*, 9-15.
2. Field, S.F., Beraldi, D., Bachman, M., **Stewart, S.K.**, Beck, S., and Balasubramanian, S. (2015). Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). PLoS ONE *10(2):* e0118202.

Acknowledgements

First and foremost, I would like to thank my supervisor, Stephan Beck, for giving me the opportunity to undertake this research and for his expert guidance at every turn. His endless patience, optimism and understanding have made all the difference to the outcome of this project and I am truly grateful. I would also like to express my sincere gratitude to everyone in the Medical Genomics lab for providing a fun and supportive working environment over the last few years. In particular, I would like to thank Andy Feber, Lee Butcher, Stefan Stricker, Helena Carén and Dirk Paul for being generous with their time and expertise through the many ups and downs of this project; this work wouldn't have been possible without their encouragement and advice.

I am immensely grateful to Gareth Wilson for performing the bioinformatics analysis of my sequencing data and to Tiffany Morris for helping me master the basics of *R*. I would also like to thank my secondary supervisor Paolo Salomoni, and his former postdocs Stefano Bartesaghi and Joanne Henderson, for introducing me to cell culture and helping me develop the neural differentiation model that is central to this thesis. Many thanks are also due to Stefan Howorka at the Department of Chemistry, along with his students Mikie Kukwikila and Hafi Nguyen, for initiating the collaboration on photocaged DNMT inhibitors and for their contributions to the project.

I am thankful to the Medical Research Council for generously supporting my research through an Industrial CASE studentship, in partnership with Diagenode. I would like to extend my thanks to the whole team at Diagenode for not only their contributions to the funding of my research, but also for hosting me at their Liège headquarters on a summer placement and for welcoming me so warmly into their R&D team. In addition, I would like to thank Toby Ost, Christine Clark and Jason Mellad at Cambridge Epigenetix for inviting us to take part in the early TrueMethyl kit trials and for facilitating helpful discussions regarding the development of the oxBS-450K method. Shankar Balasubramanian, Dario Beraldi and Sarah Field at the University of Cambridge were all instrumental in the early discussions of the oxBS-

450K method, particularly concerning data interpretation and analysis. I would also like to thank Martin Bachman, a PhD student in the Balasubramanian lab, for lending his expertise in LC-MS analysis to this project.

Finally, a huge thank you to my fellow PhD students, Paul Guilhamon, Anna Koeferle, Dimitra Georgopoulou, Hafi Nguyen, Anna Karpathakis and Tosin Taiwo, for their friendship and good humour; UCL would have been a lonely place without them. Above all, I would like to thank my parents for their never-ending love, support and encouragement. Amo vocês, obrigada por tudo que vocês fizeram por mim.

Abstract

5'-methylcytosine (5mC) plays a crucial role in the epigenetic regulation of gene expression and, until recently, was the only known epigenetic mark to result from the chemical modification of bases in mammalian deoxyribonucleic acid (DNA). The discovery of 5'-hydroxymethylcytosine (5hmC) at physiologically significant levels in a wide range of tissues, particularly those of the central nervous system, suggests that this novel epigenetic modification may have a similarly important function to 5mC in transcriptional regulation. The highest levels of 5hmC have been consistently found in fully differentiated cell types, whilst stem cells seem to be characterised by very low or insignificant levels of 5hmC. It therefore appears that loss of pluripotency is associated with a substantial increase in global 5hmC levels and this modification may play a crucial role in this switch in cell fate. The central aim of this project was to investigate the potential role of 5hmC by profiling both 5mC and 5hmC in parallel during the differentiation of embryonic stem cells (ESCs) down a neural lineage, allowing for a deeper understanding of the potential function of 5hmC in the genome. Analysis of genome-wide 5mC and 5hmC patterns in ESCs, neural stem cells (NSCs) and astrocytes supported the hypothesis that dynamic changes in the distribution of both modifications contribute to neural specification. Striking differences in 5hmC levels between *in vitro*- and *in vivo*-derived samples were observed, suggesting that cell culture models may not successfully recapitulate 5hmC profiles observed during normal development. Finally, a novel method was successfully developed and validated for genome-wide 5hmC profiling (oxBS-450K), allowing sensitive and reproducible detection of 5hmC at single-base resolution.

Abbreviations

5-aza = 5’-azacytidine

5caC = 5’-carboxycytosine

5fC = 5’-formylcytosine

5gmC = 5’-glucosyl-hydroxymethylcytosine

5hmC = 5’-hydroxymethylcytosine

5mC = 5’-methylcytosine

ACN = acetonitrile

AML = acute myeloid leukaemia

ß-GT = beta-glucosyltransferase

BER = base-excision repair

BS = sodium bisulphite

BSA = bovine serum albumin

BS-pyroseq = bisulphite pyrosequencing

BWA = Burrows-Wheeler alignment

Caged-aza = photocaged 5-azacytidine

Caged-DAC = photocaged 5-aza-2’-deoxycytidine

CGI = CpG island

CMS = cytosine-5-methylenesulfonate

CNS = central nervous system

CpG = cytosine-guanine dinucleotide

DAC = 5-aza-2’-deoxycytidine

DAPI = 4',6-diamidino-2-phenylindole

dC = deoxycytidine

dG = deoxyguanosine

DMR = differentially methylated regions

DMSO = dimethyl sulphoxide

DNA = deoxyribonucleic acid

DNMT = DNA methyltransferase

DPBS = Dulbecco’s phosphate-buffered saline

EB = embryoid bodies

EdU = 5-ethynyl-2'-deoxyuridine

EGF = epidermal growth factor

ELISA = enzyme-linked immunosorbent assay

ESC = embryonic stem cell

FBS = foetal bovine serum

FGF-2 = fibroblast growth factor 2

GFAP = glial fibrillary acidic protein

GFP = green fluorescent protein

GLIB = glucosylation, periodate oxidation and biotinylation

HAT = histone acetyltransferase

HDAC = histone deacetylase

hDMR = differentially hydroxymethylated region

hMeDIP = hydroxymethylated DNA immunoprecipitation

HPLC-MS = high-performance liquid chromatography-mass spectrometry

ICC = immunocytochemistry

LC-MS = liquid chromatography-mass spectrometry

LIF = leukaemia inhibitory factor

MACS = model-based analysis of ChIP-seq

MECP2 = methyl CpG binding protein 2

MBD = methyl-binding domain

MDS = myelodysplastic syndrome

MeDIP = methylated DNA immunoprecipitation

MeDUSA = methylated DNA utility for sequence analysis

mRNA = messenger RNA

NSC = neural stem cell

NPC = neural progenitor cell

OCT4 = octamer-binding transcription factor 4

oxBS = oxidative bisulphite

oxBS-pyroseq = oxidative bisulphite pyrosequencing

PBST = DPBS with Triton

PCNA = proliferating cell nuclear antigen

PCR = Polymerase chain reaction

PFA = paraformaldehyde

PGC = primordial germ cell

QC1/2 = quality control 1/2

qPCR = quantitative PCR

RNA = ribonucleic acid

rRNA = ribosomal RNA

SAMtools = sequence alignment/map tools

Saos-2 = sarcoma osteogenic 2

SVZ = subventricular zone

TDG = thymine DNA glycosylase

TET = ten-eleven translocation

tRNA = transfer RNA

UHRF1 = ubiquitin-like, with PHD and RING finger domains, 1

UV = ultraviolet

WGA = whole genome amplification

WGBS = whole genome bisulphite sequencing

Chapter 1 Introduction

1.1 The epigenetic code

Normal developmental processes are dependent on stable, long-term modifications to the transcriptional potential of a cell that do not affect the underlying nucleotide sequence. These epigenetic modifications function alongside the genetic code to shape cell phenotype during differentiation through the control of chromatin dynamics and gene expression patterns. Mechanisms of epigenetic activity typically refer to histone modifications, non-coding ribonucleic acid (RNA) signalling and the methylation of cytosine bases within DNA, which together regulate the transcriptional profile of a cell (Figure 1-1). Epigenetic modifications may be mitotically inherited across multiple cell divisions and thus contribute to stable cellular identity, whilst DNA methylation has also been shown to exhibit transgenerational heritability (Chong and Whitelaw, 2004; Holliday, 1987; Margueron and Reinberg, 2010).

Epigenetic states are considered dynamic sets of modifications that may undergo changes during cellular reprogramming events or in response to environmental cues (Feil and Fraga, 2012; Meissner, 2010). Site-specific, temporal changes are observed in the 5'-methylcytosine (5mC) content of gene promoters, whilst two waves of global DNA demethylation take place during early embryogenesis, first in the paternal pronucleus at the one-cell stage of zygotic development, and then again in the primordial germ cells (PGCs) of the developing embryo (Kangaspeska et al., 2008; Mayer et al., 2000; Metivier et al., 2008; Oswald et al., 2000; Santos et al., 2002). Histone modifications also undergo global changes during development; during spermatogenesis, the majority of histone proteins are replaced by protamines throughout the genome and then restored after fertilisation (Hammoud et al., 2009; Torres-Padilla et al., 2006).

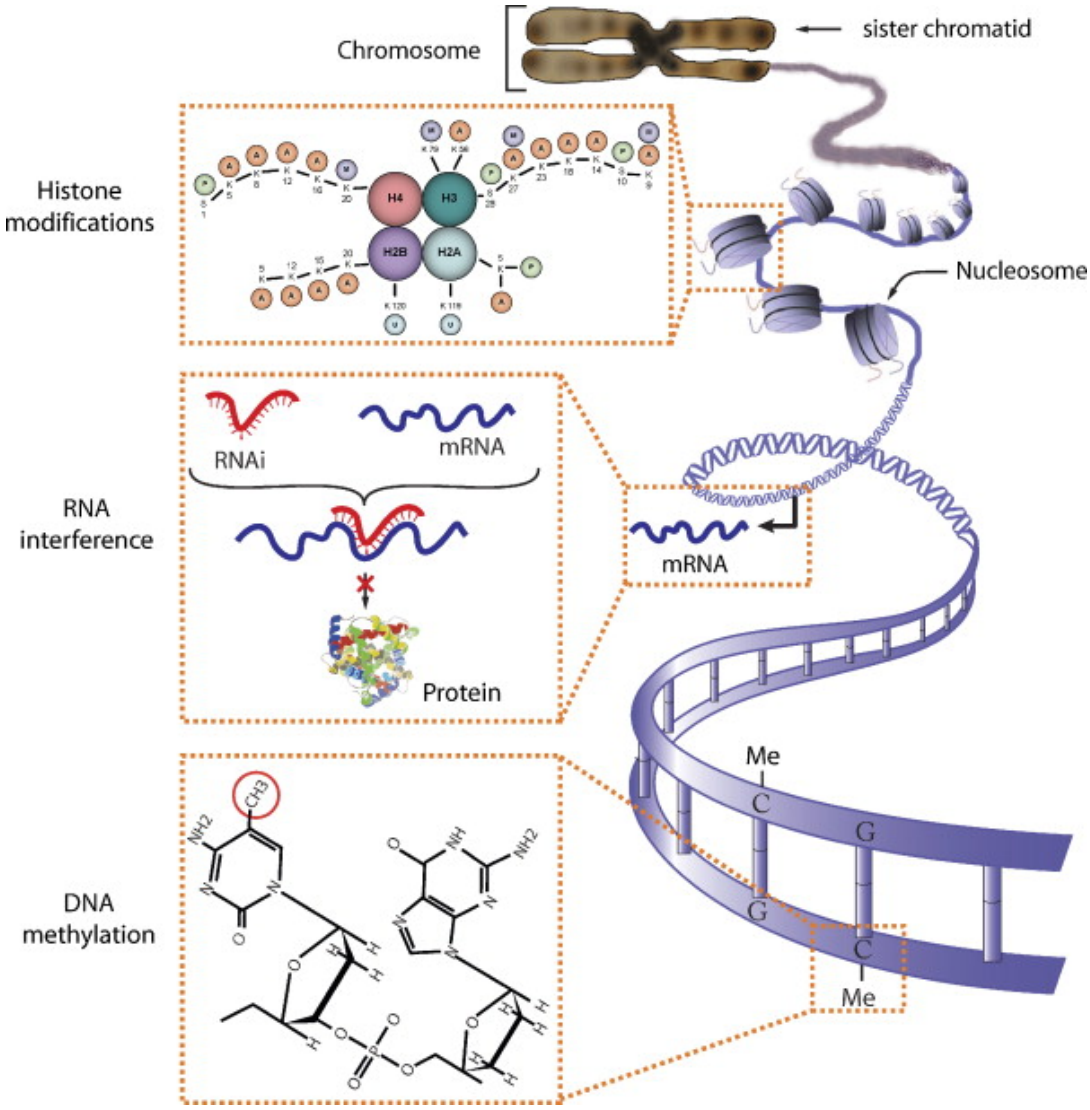


Figure 1-1: Overview of epigenetic modifications within the mammalian genome. Non-coding RNA molecules, chemical modifications of histone proteins within nucleosomes and methylation of cytosine bases all contribute to the layer of transcriptional control over the genome known as the epigenome. Image courtesy of: Sawan et al., *Mutation Research* **642 (1-2),** 1-13 (2008).

1.2 DNA methylation

1.2.1 5'-methylcytosine

Methylation at the carbon-5 (5') position of cytosine is a post-replication modification of DNA present in most eukaryotes, including all vertebrates and plant species, and many prokaryotes (Goll and Bestor, 2005; Suzuki and Bird, 2008). DNA methylation is critical for normal development and plays a role in an array of cellular processes such as chromatin remodelling and transcriptional regulation, X chromosome inactivation, genomic imprinting and transposon silencing (Bird, 2007; Kiefer, 2007). 5mC is strongly associated with heterochromatic regions of the genome where it functions as a permanent, silencing epigenetic modification (Hashimshony et al., 2003; Jaenisch, 1997; Walsh et al., 1998). The localisation of 5mC at promoters usually results in transcriptional silencing and it has therefore traditionally been regarded as a repressive epigenetic mark (Cedar and Bergman, 2009; Siegfried et al., 1999; Tiwari et al., 2008; Zhang et al., 2005), although DNA methylation is frequently correlated with active coding regions and tissue-specific gene expression (Futscher et al., 2002; Grunau et al., 2000; Kumar and Biswas, 1988; Ngô et al., 1996; Song et al., 2005). For example, 5mC is found in the gene body of actively transcribed genes in both plants and mammals (Ball et al., 2009; Hellman and Chess, 2007; Zhang et al., 2006).

DNA methylation is arguably the best understood and most widely studied epigenetic modification in mammalian cells; the mechanisms controlling the establishment and maintenance of DNA methylation patterns are well characterised and functional studies have shown that DNA methylation is essential for normal development and cellular differentiation (Jackson et al., 2004). The DNA methyltransferase (DNMT) family of enzymes are responsible for the establishment and maintenance of 5mC modifications in the genome (Bestor, 2000; Goll and Bestor, 2005). DNMT3A and DNMT3B are known as the *de novo* methyltransferases and establish DNA methylation patterns during gametogenesis and early embryonic development (Okano et al., 1999), whilst DNMT1 has been identified as a maintenance methyltransferase that conserves DNA methylation patterns on daughter DNA strands during replication (Lei et al., 1996; Li et al., 1992). DNMT1 associates with the replication fork during DNA synthesis through

binding to PCNA and UHRF1 and preferentially targets hemi-methylated DNA sites to restore 5mC on daughter strands; loss of either DNMT1 or UHRF1 triggers passive DNA demethylation (Bostick et al., 2007; Chuang et al., 1997; Sharif et al., 2007). The two remaining DNMT enzymes, DNMT3L and DNMT2, retain significant structural homology to the other DNMT enzymes but contain no detectable DNA methylation activity (Yoder and Bestor, 1998). DNMT2, now more commonly known as tRNA aspartic acid methyltransferase 1 (TRDMT1), instead functions as an RNA methyltransferase (Goll et al., 2006), whilst DNMT3L acts as a regulator of *de novo* methylation through its interaction with DNMT3A and DNMT3B (Chen et al., 2005; Jia et al., 2007; Suetake et al., 2004). DNMT3L is essential for the establishment of mammalian germline imprinting and is thought to directly recruit DNMT3A2 (an isoform of DNMT3A present in germ cells) to DNA through its specific binding to the unmethylated histone H3K4 (Bourc'his et al., 2001; Hata et al., 2002; Ooi et al., 2007).

1.2.2 5'-hydroxymethylcytosine and further oxidative derivatives

More recently, the ten-eleven translocation (TET) proteins have been identified as 2-oxoglutarate- and iron-dependent dioxygenases that are capable of converting 5mC to its further oxidation products: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009).

5hmC was first identified at physiologically relevant levels in mouse ESCs, Purkinje neurons, and granule cells in the brain, with levels ranging from 0.03% (in ESCs) to 0.6% (in Purkinje cells) of total nucleotides, up to 40% as abundant as 5mC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). 5hmC has since been identified in a wide variety of mammalian tissues and cell lines, suggesting that 5hmC may be a biologically relevant epigenetic modification in its own right with unique gene regulatory functions. However, the highest enrichment levels are consistently found in tissues of the CNS, where 5hmC accounts for approximately 0.3 - 0.7% of the total cytosine content of the cells (Globisch et al., 2010a; Li and Liu, 2011; Song et al., 2011b; Szwagierczak et al., 2010). 5hmC levels in adult mNSCs have been found to be similar to that of mESCs, at around 0.04% of total

nucleotides, whilst an age-dependent increase in global 5hmC is observed in mouse cerebellum. 5hmC comprises 0.1% of total nucleotides at postnatal day 7, reaching a maximum of 0.4% at 2.5 months old (Song et al., 2011b).

5fC and 5caC have also been detected within mammalian DNA, but at levels at least 10-fold lower than that of 5hmC. A sensitive high-performance liquid chromatography-mass spectrometry (HPLC-MS) method was able to detect 5fC, but not 5caC, in mESC DNA at levels of 0.02% of the total deoxyguanosine (dG) content, around 20-fold lower than 5hmC (0.39%/dG) and over 200-fold lower than the 5mC content (5%/dG) of the DNA sample (Pfaffeneder et al., 2011). Another study utilising an HPLC-MS assay was able to quantify both the 5fC and 5caC content of mESC DNA (Ito et al., 2011). 5caC was present at a level of 3 5caC per 10^6 C bases, approximately 7-fold lower than 5fC and over 400-fold lower than the 5hmC content of the DNA sample. Nuclear extracts from mESCs have been shown to contain a specific thymine DNA glycosylase (TDG) activity that is able to excise 5fC and 5caC from DNA substrates, but not 5hmC, 5mC or unmodified C, perhaps explaining why neither 5fC or 5caC appear to accumulate to significant levels within genomic DNA (He et al., 2011; Zhang et al., 2012).

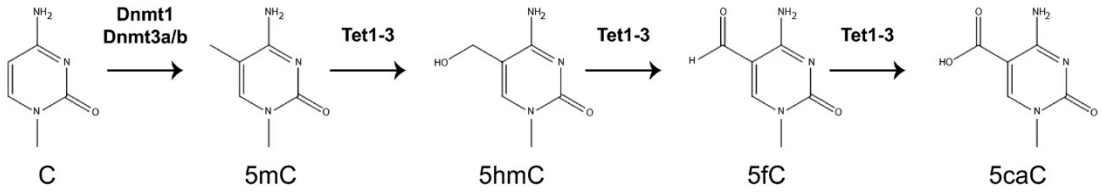


Figure 1-2: The chemical structure of cytosine modifications present in mammalian DNA. A methyl (CH_3) group may be added to the 5' position of the cytosine pyrimidine ring by the DNMT enzymes, which can then undergo iterative oxidation reactions by the TET enzymes to produce a primary alcohol (5hmC), aldehyde (5fC) or carboxylic acid (5caC). Image courtesy of: Hill et al., *Genomics* **104 (5),** 324-333 (2014).

1.2.3 The genomic context of 5mC

In mammalian somatic tissue, over 99% of all 5'-methylcytosine (5mC) bases occur in the context of CpG dinucleotides, although a significant proportion of non-CpG methylation has been found within ESCs and adult neurons, making up around 25% and 53% of all 5mC, respectively (Bernstein et al., 2007; Lister et al., 2009, 2013). CpG sites make up approximately 1-4% of mammalian DNA and around 70% are

methylated (Bird et al., 1985; Ehrlich et al., 1982; Lister et al., 2011). CpG sites are generally depleted within the genome due to the susceptibility of 5mC to mutation via deamination to produce thymine, but certain areas of high CpG density, known as CpG islands (CGIs), remain protected as they are largely unmethylated in the germ line (Barker et al., 1984; Bird, 1980). Although definitions vary, CGIs are usually characterised by a minimum GC content of 50-55%, a minimum observed:expected CpG ratio of 0.60-0.65 and a defined sequence length, often between 200-1000bp (Bird, 1986; Gardiner-Garden and Frommer, 1987; Illingworth and Bird, 2009; Takai and Jones, 2002). It has been estimated that there are over 25,000 CGIs in the human genome and around 23,000 CGIs in the mouse genome, although these regions show a lower average CpG density in mouse (Illingworth et al., 2010). Over 50% of gene promoter regions are associated with CGIs (Antequera and Bird, 1993; Bird, 1987) and methylation events at these CGIs are highly correlated with chromatin accessibility and gene transcription states, indicating a regulatory role for 5mC at these promoter-associated regions (Deaton and Bird, 2011).

Between 40-60 million 5mC bases are found within mammalian DNA, making up around 4-6% of all cytosines. 5mC is widespread in its genomic location and is found both in gene bodies and more distal regions. Approximately 80% of CpG methylation is concentrated in repeat DNA sequences such as satellites, whilst CGIs and transcriptional start sites (TSS) are frequently depleted of 5mC (Bird, 1986; Ehrlich et al., 1982). Within intragenic regions, similar levels of relatively high methylated CpG density are found in introns, exons and 3' UTR regions, with levels decreasing in the 5' UTR region as you move nearer the TSS, which shows minimal methylation. Levels of 5mC then increase again as you move upstream of the TSS, returning to similar gene body methylation levels around 2kb upstream of the TSS. Non-CpG methylation within ESCs is also enriched in the gene body and depleted at the TSS, with 15-20% higher 5mC levels observed in exons over introns and the 3'UTR. Unlike CpG methylation, non-CpG 5mC levels in ESCs are around 50% lower in the regions flanking the TSS (5' UTR and promoter) than in the gene body. However, both CpG and non-CpG methylation are relatively depleted at transcription factor (TF) binding sites and active enhancer regions. Non-CpG methylation density in the gene body is positively correlated to gene expression, whilst no correlation is

detected with CpG methylation density (Lister et al., 2009). This is in contrast with gene body CpG methylation in differentiated somatic cells, where a positive correlation to gene expression levels has been observed in human fibroblast and B-lymphocyte cell lines (Ball et al., 2009; Lister et al., 2009).

1.2.4 The genomic context of 5hmC

Initial reports of high levels of 5hmC in ESCs (compared to negligible levels in most somatic tissues) led to speculation that 5hmC may be responsible for the high proportion of non-CpG methylation detected in ESCs through bisulphite-based assays. Some of the first genome-wide 5hmC profiles generated by hMeDIP-seq reported extensive enrichment of strand-specific 5hmC in CAG sequences, as well as CA and CT repeat regions (Ficz et al., 2011; Matarese et al., 2011; Williams et al., 2011). However, the development of single-base resolution methods of 5hmC detection have revealed that around 99% of 5hmC is found in the CpG context (both in mouse and human ESCs and brain tissue), suggesting that previous reports of non-CpG 5hmC were probably due to technical limitations and/or bias of the affinity enrichment assays used (Booth et al., 2012; Lister et al., 2013; Wen et al., 2014; Yu et al., 2012). Nevertheless, contradictions remain in the current body of literature concerning genomic 5hmC distribution, particularly regarding suggestions of strand asymmetry, repeat element enrichment and the precise relationship between 5mC and 5hmC coexisting at the same loci. Further technological advances in 5hmC detection methods will no doubt be needed to resolve these ambiguities.

The first genome-wide maps of 5hmC in both mouse ESC and cerebellum DNA, using affinity enrichment techniques coupled to sequencing, found enrichment concentrated in intragenic regions, with little or no detectable 5hmC in repeat regions (Pastor et al., 2011; Song et al., 2011b; Williams et al., 2011; Wu et al., 2011). Particularly high 5hmC enrichment within exons was also observed in ESCs, as has previously been observed for 5mC (Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011). Indeed, the presence of 5hmC appears to require pre-existing 5mC, after it was shown that triple DMNT knockout cells lack both 5mC and 5hmC in the genome (Ficz et al., 2011; Williams et al., 2011). 5hmC levels within the gene body show a positive correlation to gene expression, suggesting a possible role for 5hmC in

promoting transcriptional activity, perhaps counteracting the repressive action of 5mC without needing to demethylate (Song et al., 2011b; Wu et al., 2011). Although the majority of 5hmC-enriched regions were also enriched for CpG sequences, a relatively high level of strand-biased CpH and CAG sequence were also identified in 5hmC enrichment profiles, suggesting that 5hmC may be present in a strand-specific, non-CpG context. (Ficz et al., 2011; Matarese et al., 2011; Pastor et al., 2011).

Depletion of 5hmC around the TSS, transcriptional termination sites (TTS) and distal intergenic regions was observed compared to proximal upstream and downstream regions, using a glucosylation-based enrichment assay in mouse cerebellum, Purkinje cells, granule cells and Bergmann glial cell DNA (Mellen et al., 2012; Song et al., 2011b). However, enrichment for 5hmC at the TSS was seen in mESCs using both glucosylation-based and CMS-based enrichment assays (see section 1.4.2), predominantly at a subset of promoter-associated CGIs of transcriptionally inactive genes characterized by bivalent histone domains and low 5mC levels (Pastor et al., 2011; Wu et al., 2011). Similarly, two hMeDIP-seq datasets also found slight 5hmC enrichment at the TSS and promoter-associated CGIs (Ficz et al., 2011; Williams et al., 2011), with a particular enrichment of 5hmC at TET1-bound TSS: between 35-50% of all genes enriched for 5hmC at the TSS were also bound by TET1 (Williams et al., 2011). 5hmC enrichment has also been reported at H3K4me1-marked enhancers (Pastor et al., 2011), the promoters of long interspersed nuclear elements (LINE1), CTCF and pluripotency transcription factor (TF) binding sites (Ficz et al., 2011), suggesting a link between 5hmC and transcriptional regulation in ESCs.

A similar genomic 5hmC distribution was detected in human frontal lobe brain tissue using an immunoprecipitation technique coupled to NimbleGen 720k CpG island plus promoter arrays (Jin et al., 2011a). The majority of 5hmC was found at promoters (55-59%) whilst a significant proportion of 5hmC was enriched in intragenic regions (35-38%). Only 6% of total 5hmC was identified in intergenic regions, suggesting selective targeting of 5hmC to genes.

5hmC identification at single-base resolution has since been made possible due to the development of two methods of selective chemical conversion of 5hmC bases within DNA, oxidative bisulphite sequencing (oxBS-seq) and TET-assisted bisulphite

sequencing (TAB-seq), which allow quantitative 5hmC mapping (see section 1.4.3 for details). Using a reduced representation oxBS-seq method that targeted CGI-containing genomic DNA from mESCs, the highest 5hmC levels were found in both intragenic and intergenic CGIs and the lowest levels at TSS-associated CGIs, supporting previous reports of intragenic 5hmC enrichment. A total of 800 out of 12,660 CGIs covered by the sequencing data (averaging 120 reads per CpG) were found to contain 5hmC, with an average level of 3.3% (Booth et al., 2012). Within CGIs, 5hmC was identified only within the CpG context and both 5mC and 5hmC levels across all CGIs were negatively correlated to CpG density. In addition, the highest 5hmC levels were found at CGIs with intermediate levels of 5mC (25-75%), and significant amounts of 5hmC were also observed at LINE1 elements, as noted by Ficz et al. (2011).

Overall, these results were confirmed by 5hmC maps generated by TAB-seq for both human H1 ESCs and mouse E14 ESCs, although the sensitivity of this method was slighter lower due to a reduced sequencing depth of 26.5 (Yu et al., 2012). 5hmC was found predominantly at CpG dinucleotides, with 99.9% of 5hmC identified in the CpG context in H1 cells and 98.7% in E14 cells, and displayed greatest enrichment in regions with the lowest CpG density. Within genic regions, 5hmC enrichment was inversely correlated to expression levels, as previously noted (Booth et al., 2012; Pastor et al., 2011). Analysis of 5hmC distribution within genomic features, however, revealed differing patterns of enrichment from previous studies. Approximately half of all 5hmC bases were located at distal-regulatory elements, including p300 and CTCF binding sites, enhancers and DNase1 hypersensitive sites. The highest absolute levels of 5hmC were found at distal-regulatory elements, compared to either proximal-promoter regions or intragenic regions, which showed the lowest 5hmC levels. However, the number of 5hmC bases in genic regions may have been underestimated due to insufficient sequencing depth, due to the very low 5hmC levels within genic regions (3-4%). An inverse relationship was noted between 5mC and 5hmC content, indicating that the strongest DNA demethylation is occurring at distal regulatory sites that show high 5hmC levels and reciprocally low 5mC.

The median abundance of 5hmC at enriched sites was measured at 20.9%, compared to a median of 59.0% for 5mC. Interestingly, 5hmC was found to show extensive strand asymmetry, unlike 5mC (Yu et al., 2012). Only 21.0% of 5hmCs were located within symmetrically modified CpG sites, whilst 91.8% of 5mC were symmetrically distributed, supporting earlier findings from hMeDIP-seq profiles (Ficz et al., 2011). In addition, low but significant enrichment of 5hmC was found in multiple repeat elements, including short interspersed nuclear elements (SINE), LINE1, and long terminal repeats (LTRs), with 5hmC levels falling between 3-12% and comparable to that seen in promoter regions.

1.2.5 Proposed mechanisms of DNA demethylating activity

Dynamic changes in DNA methylation patterns are frequently observed during development and are important determinants in cell fate specification and cellular reprogramming in germ cells (Liang et al., 2011; Meissner et al., 2008a; Suzuki and Bird, 2008). Evidence for both locus-specific and genome-wide DNA demethylation events have originated from studies of early embryogenesis. In the fertilised oocyte, the paternal pronucleus undergoes a global wave of DNA demethylation prior to pronuclear fusion, suggesting the existence of an active mechanism of 5mC erasure that is independent of DNA replication (Hajkova et al., 2002; Oswald et al., 2000; Santos et al., 2002). The maternal pronucleus also undergoes extensive global DNA demethylation, albeit more slowly during the first few cleavage divisions of the early embryo, due to a replication-dependent, passive dilution of 5mC (Rougier et al., 1998).

The recent identification of 5-hydroxymethylcytosine and further oxidation products of 5-methylcytosine has led to the suggestion that they represent intermediates of a DNA demethylation mechanism initiated by the TET enzyme family (Bhutani et al., 2011; Wu and Zhang, 2010). DNMT1 shows a reduced binding preference for 5hmC compared to 5mC, suggesting that oxidation of 5mC to 5hmC by the TET proteins may be a method of disrupting the maintenance methylation machinery and triggering passive demethylation through the dilution of 5mC during successive rounds of DNA replication (Tahiliani et al., 2009). 5hmC immunostaining of mitotic

chromosomes in pre-implantation embryos provided further support of such a replication-dependent, passive demethylation mechanism (Inoue and Zhang, 2011).

However, the precise mechanisms governing active DNA demethylation have been more difficult to elucidate. No mammalian proteins have so far been identified with 5mC-specific DNA base-excision repair demethylase activity analogous to the DNA glycosylases found in plants, but the base-excision repair (BER) pathway has been shown to play an essential role during *in vivo* active DNA demethylation in mammals (Hajkova et al., 2010). Thymine DNA glycosylase (TDG)-deficient embryos are non-viable and characterised by severe epigenetic abnormalities, highlighting the importance of TDG in normal development (Cortázar et al., 2011; Cortellino et al., 2011). Two main pathways have been proposed to modify 5mC to a suitable TDG substrate: TET-mediated oxidation and deamination to thymine by the AID (activation induced cytidine deaminase) or APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) proteins (Figure 1-3).

A TET- and TDG-mediated mechanism was proposed after the discovery of the TET family of enzymes in mammalian cells and their ability to catalyse the sequential oxidation of 5mC in both the CpG and non-CpG context *in vitro* to 5hmC, 5fC and 5caC, followed by thymine DNA glycosylase-mediated excision repair (Guo et al., 2011; Ito et al., 2011; Tahiliani et al., 2009). TDG specifically recognizes both 5fC and 5caC bases (but neither 5mC or 5hmC) within duplex DNA, suggesting a mechanistic link for both TET and TDG to active DNA demethylation (He et al., 2011; Maiti and Drohat, 2011; Zhang et al., 2012). The preference of TDG for removing 5fC and 5caC from C:G pairings, despite the greatly reduced numbers of 5fC/5caC within the genome compared to unmodified or methylated C, was explained though the observation that TDG base-excision activity is dependant on N-glycosidic bond stability and that 5fC/5caC have the lowest energy barriers for N-glycosidic bond cleavage than either 5hmC, 5mC or unmodified C (Bennett et al., 2006; Williams and Wang, 2012). Several other observations have strengthened the view that TET/TDG-mediated excision represents the most likely mechanism behind active DNA demethylation: TET3 knockdown results in only partial demethylation of the paternal pronucleus in fertilised oocytes and eventual embryonic lethality (Gu

et al., 2011); TDG knockdown in mESCs results in an accumulation of global 5caC levels, whilst TDG overexpression reduces 5caC (He et al., 2011); and TET1 activity is essential for the erasure of paternal genomic imprints in PGCs (Yamaguchi et al., 2013).

It has also been suggested that the oxidised methyl group could be directly removed by a 5caC-specific decarboxylase, although the existence of such an enzyme remains unknown (Schiesser et al., 2012). There is some evidence for an alternative mechanism of active DNA demethylation involving TDG and the AID/APOBEC enzyme complex, whereby either 5mC or 5hmC undergo deamination to thymine and 5-hydroxymethyluracil, respectively, followed by TDG-mediated BER.

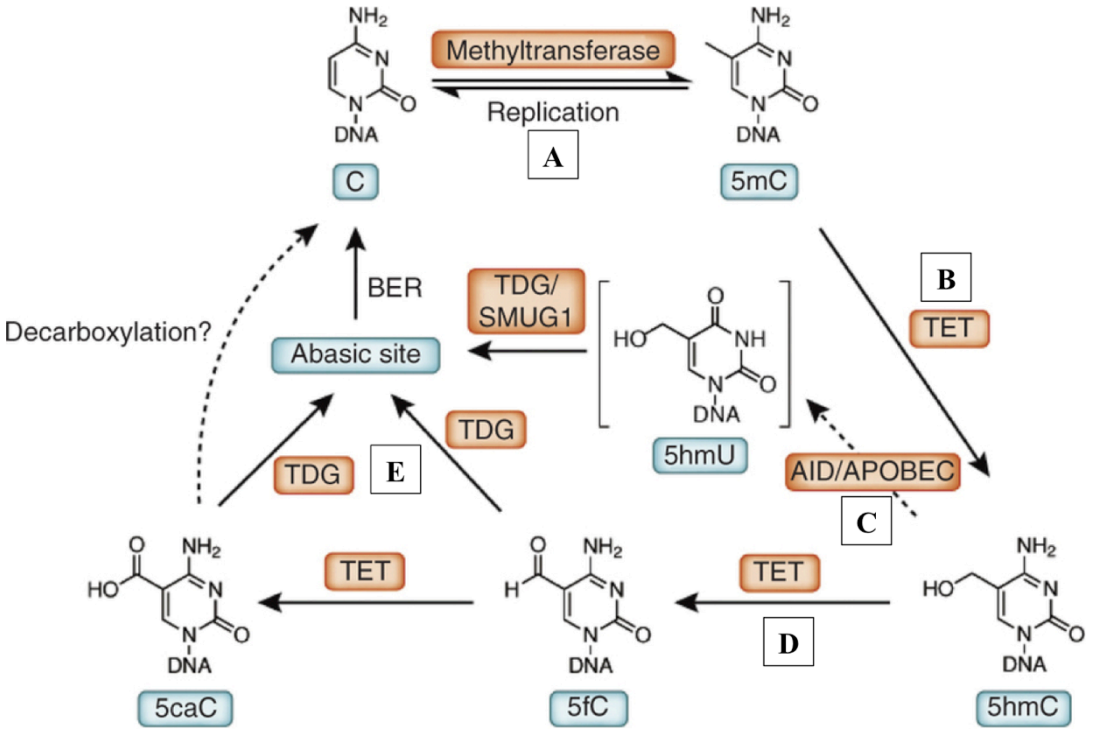


Figure 1-3: Schematic of proposed DNA demethylation pathways and their enzymatic components. DNA demethylation may occur either through the passive dilution of 5mC through replication (A), or through a replication-independent active mechanism involving TET oxidation to 5hmC (B). 5hmC may then undergo either deamination by the AID/APOBEC enzymes to 5hmU (C) or further oxidation to 5fC/5cacC (D). TDG-mediated BER pathway can target either 5fC, 5caCor 5hmU substrates and restore an unmodified C:G pairing in duplex DNA (E). Image adapted from: Song et al., *Nature Biotechnology* **30(11),** 1107-1116 (2012).

TDG is known to recognise both thymine (T)-G and uracil (U)-G mismatches within DNA and excise the mismatched base, thus restoring normal C-G base pairing (Neddermann and Jiricny, 1993, 1994). It has been proposed that 5mC deamination

by the AID/APOBEC enzyme complex to T, followed by TDG-mediated excision repair, could be a method of active DNA demethylation. The AID enzyme appears to be required for active DNA demethylation during both experimental cellular reprogramming of somatic cells and during *in vivo* reprogramming of PGCs (Bhutani et al., 2010; Popp et al., 2010). TDG is also able to recognize and excise 5-hydroxymethyluracil (5hmU)-G mismatches, suggesting that deamination of 5hmC could be a possible demethylation mechanism (Cortellino et al., 2011; Guo et al., 2011). However, 5hmC deamination has not so far been detected either *in vivo* or *in vitro* by the AID/APOBEC deaminases, which also show a much reduced preference for 5mC substrates over unmodified C (Nabel et al., 2012; Rangam et al., 2012). Moreover, AID deficiency does not result in embryonic lethality, unlike TDG (Popp et al., 2010). Given the critical role of 5mC in development, active DNA demethylation must entail an alternative AID-independent mechanism.

1.3 5mC profiling methods

1.3.1 Restriction enzyme-based assays

Changes in DNA methylation patterns have been implicated in a wide variety of developmental and disease processes, driving the rapid development of novel methods of methylome analysis (Beck and Rakyan, 2008; Brena et al., 2006). DNA methylation profiling typically relies on an initial 5mC-dependent treatment, such as restriction enzyme digestion, affinity enrichment or bisulphite conversion, coupled to a detection method such as PCR amplification, microarray hybridisation or DNA sequencing (Laird, 2010).

The first example of DNA methylation detection in the late 1970s used methylation-sensitive restriction enzyme digestion of DNA, followed by analysis using gel electrophoresis and Southern blotting, to show that methylation in mammalian DNA occurs predominantly in the CpG context (Gautier et al., 1977). Commonly used restriction enzymes to interrogate 5mC include HpaII and MspI, a pair of isoschizomers with the same recognition site CCGG but differential methylation sensitivity: HpaII is inhibited by methylation of the second cytosine in the recognition sequence ($C^{me}CGG$), whereas MspI is not. DNA cleavage with HpaII and MspI in separate assays and comparison of the resulting fragment sizes allows

for the discrimination of methylated from unmethylated DNA. Locus-specific, quantitative 5mC detection may also be achieved by combining restriction enzyme digestion with PCR across the recognition sequence, although this method is prone to false positive readings due to incomplete HpaII digestion (Singer-Sam et al., 1990). Nevertheless, differential enzymatic 5mC sensitivity forms the basis of more recently developed genome-wide detection methods such as the HELP assay (HpaII tiny fragment enrichment by ligation-mediated PCR) and Methyl-Seq, which combine restriction enzyme digestion of genomic DNA with microarray hybridisation and next-generation sequencing, respectively, to identify methylated loci (Brunner et al., 2009; Khulan et al., 2006).

1.3.2 Affinity enrichment-based assays

Affinity enrichment techniques use either antibodies raised against 5mC or the methyl-binding domain (MBD) of MECP2, a methylation-specific CpG-binding protein, to efficiently precipitate 5mC-containing DNA fragments. The first demonstration of a 5mC affinity enrichment assay used an MBD column (containing a MECP2-bound nickel-agarose matrix) to fractionate human genomic DNA according to its methylated CpG content (Cross et al., 1994). Alternative MBD proteins with higher affinity for methylated CpG sites were subsequently engineered (Jørgensen et al., 2006) and used for profiling captured DNA on microarrays (Keshet et al., 2006). MBD enrichment may be coupled with PCR for targeted, locus-specific 5mC detection or coupled with array hybridisation for genome-wide analysis (Gebhard et al., 2006a, 2006b). The methylated CpG island recovery assay (MIRA) is a similar technique for MBD-mediated enrichment (using an MBD2/MSB3L1 complex) that may be coupled to arrays or next-generation sequencing for 5mC analysis (Rauch and Pfeifer, 2005; Rauch et al., 2006, 2007).

Enrichment using an anti-5mC antibody, termed methylated DNA immunoprecipitation (MeDIP), was used to produce the first genome-wide 5mC profiles in mammalian DNA in combination with tiling microarray hybridisation (Weber et al., 2005). The advent of next generation, high-throughput DNA sequencing has allowed for genome-wide 5mC profiling using both MeDIP and MBD enrichment strategies without being restricted by either the species-specificity

of a microarray or to the small subset of genomic loci typically represented (Brinkman et al., 2010; Down et al., 2008; Serre et al., 2009). MeDIP-seq has become a well-established method of generating methylomes in a efficient, cost-effective manner that in amenable to high-throughput sample processing and automation of the experimental protocol (Butcher and Beck, 2010). Both MeDIP-seq and MBD-seq offer moderate genomic resolution of 100-300bp, depending on DNA fragment size and sequencing read length, and can accurately identify differentially methylated regions (DMRs) between samples, but may only be used to infer relative methylation differences rather than quantitative differences because the detected 5mC signals are strongly influenced by sequencing depth. In addition, both enrichment methods exhibit bias towards CpG-dense regions of the genome, potentially underestimating the proportion of 5mC in areas of low CpG content (Bock et al., 2010; Harris et al., 2010; Li et al., 2010a).

1.3.3 Sodium bisulphite-based assays

Sodium bisulphite sequencing (BS-seq) is considered the gold standard for genome-wide DNA methylation profiling, combining BS conversion of DNA with next-generating sequencing technology to produce single-base resolution DNA methylomes (Cokus et al., 2008; Frommer et al., 1992). BS conversion can differentiate methylated from unmethylated cytosines in the genome but does not discriminate between 5mC and 5hmC, which are protected from conversion to uracil and are both read as cytosine after sequencing (Huang et al., 2010; Jin et al., 2010; Nestor et al., 2010). The first single-base-resolution, mammalian DNA methylome was generated using BS-seq for human ESCs and foetal fibroblasts (Lister et al., 2009) and similar studies have since employed BS-seq to profile 5mC in a wise variety of samples, including human cancers, cord blood, peripheral blood mononuclear cells and mouse frontal cortex (Berman et al., 2011; Heyn et al., 2012; Hon et al., 2012; Li et al., 2010b; Xie et al., 2012). However, BS-seq requires much deeper sequencing to generate 5mC profiles with comparable CpG coverage to those obtained through affinity-enrichment methods, substantially increasing the cost per methylome. Whilst assays such as MeDIP- and MBD-seq produce DNA libraries covering only highly-methylated genomic regions, thus concentrating the sequencing effort only on potential regions of interest, BS-seq datasets cover the entire genome

and the majority of sequencing reads end up mapping to unmethylated genomic regions and are effectively discarded (Robinson et al., 2010).

Reduced representation bisulphite sequencing (RRBS) employs a similar experimental method to profile DNA methylation in a defined subset of CG-rich genomic regions. RRBS covers less than 10% of all CpG sites and is typically restricted to CpG islands and promoter regions, but allows for a much more cost-effective genome-wide 5mC analysis than whole genome BS-seq, making it suitable for profiling multiple DNA samples (Bock et al., 2010; Gu et al., 2010; Meissner et al., 2005, 2008b).

BS-converted DNA is not easily interrogated through standard array hybridization techniques in the same way as affinity-enriched or endonuclease-digested DNA because the lower sequence complexity results in frequent mismatches and decreased binding specificity. However, several specialised arrays have been developed specifically for the detection of genome-wide DNA methylation at single-base resolution. The Infinium HumanMethylation450 (450K) BeadChip provides the most comprehensive genome coverage of the currently available arrays, interrogating 485,577 CpG sites, 3,091 non-CpG sites and 65 random SNPs, and is arguably the most widely used DNA methylation profiling microarray. Sample preparation requires minimal hands-on time and each BeadChip consists of 12 arrays for DNA hybridisation, making this method suitable for analysing large cohorts. The probes cover 99% of all RefSeq genes and target multiple transcripts, covering a total of 29,246 transcripts with an average of 17.9 probes each (Bibikova et al., 2011). Probes were designed to target multiple gene features, including promoters, gene bodies, the first exon, 5' and 3' UTRs (Figure 1-4A). In addition, 96% of all CpG islands are covered by the array's probes, where CpG islands are defined as 500bp windows with a GC base composition over 55% and an observed:expected CpG ratio over 0.65, totalling 26,658 targeted CpG islands with an average of 5.63 probes each (Figure 1-4B).

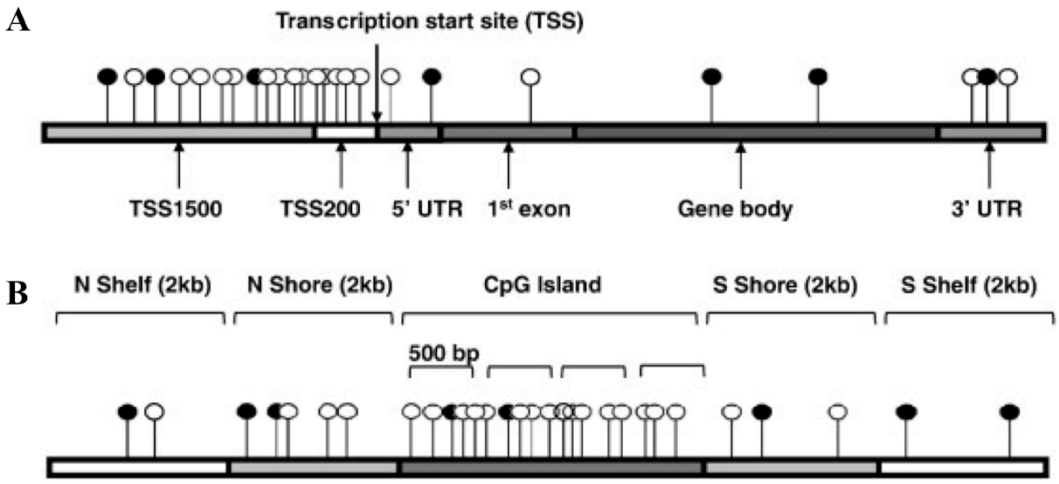


Figure 1-4: Infinium 450K array probe distribution and design. A) Typical probe representation of each gene transcript. B) Typical probe representation of regions of differential CpG context. Image courtesy of: Bibikova et al., *Genomics* **98(4),** 288-295 (2011).

The 450K platform has been extensively validated against other technologies as a highly sensitive and reproducible 5mC profiling method and is also compatible with formalin-fixed, paraffin-embedded (FFPE) tissue samples (Dedeurwaerder et al., 2011; Roessler et al., 2012; Sandoval et al., 2011). The moderate genome coverage of the Infinium 450K array compared to other sequencing-based assays, such as RRBS-seq, is counterbalanced by the substantially lower cost, higher sample throughput and less complex data analysis inherent in this assay, which has established itself as the current platform of choice for population-based epigenetic analyses, such as epigenome-wide association studies (Liu et al., 2013b; Michels et al., 2013; Rakyan et al., 2011; Tsai and Bell, 2015). However, conventional 450K protocols depend on BS conversion of DNA and therefore cannot be used to discriminate 5mC from 5hmC within the genome. The popularity of the 450K array for epigenetic studies made it an attractive candidate for adaptation to 5hmC analysis after the emergence of oxidative bisulphite (oxBS) chemistry (Booth et al., 2012). This method is explored further in Chapter 5.

For locus-specific detection, combined bisulphite restriction analysis (COBRA) relies on restriction enzyme digestion of BS-converted DNA followed by gel electrophoresis, electroblotting and oligonucleotide hybridisation to quantify 5mC at particular recognition sites (Xiong and Laird, 1997). This method is suitable for use

with FFPE tissue samples and small amounts of starting DNA, but is labour-intensive and limited to assaying a defined set of restriction sites. Similarly, methylation-specific PCR (MSP) of BS-converted DNA using methylation-sensitive primers followed by gel electrophoresis may be used for semi-quantitative assessment of 5mC levels and is sensitive enough to detect very low 5mC levels (0.1%), but it is also a fairly labour-intensive method and therefore not suitable for large-scale analyses (Herman et al., 1996). For high-throughput, quantitative PCR-based 5mC detection, MethyLight assays can rapidly and accurately identify methylated DNA loci using real-time, fluorescence-based TaqMan® technology (Eads et al., 2000). Locus-specific BS-seq can be used to interrogate 5mC by sequencing PCR products amplified from BS-converted DNA, allowing strand-specific, single-base resolution of 5mC. However, this method but requires cloning of PCR amplicons followed by individual sequencing of multiple clones in order to provide adequate sensitivity towards 5mC (Frommer et al., 1992).

Pyrosequencing technology allows accurate, quantitative measurement of 5mC at individual CpG sites in a strand-specific manner (Dejeux et al., 2009; Ronaghi et al., 1996; Tost and Gut, 2007). Pyrosequencing analysis of DNA methylation relies on BS conversion of DNA samples followed by PCR amplification of the target sequence using biotin-labelled primer pairs. These amplicons are then subjected to a real-time DNA sequencing-by-synthesis method, which relies on a cascade of enzymatic reactions following incorporation of a nucleotide in the newly synthesised DNA strand to generate visible light in a manner proportional to the number of added nucleotides. Only one type of nucleotide is present in the reaction well at any one time and unincorporated nucleotides are degraded rapidly, allowing consecutive addition of the four nucleotides to the reaction wells. The intensity of the light generated for each nucleotide type is measured and used to determine the sequence of the template DNA strand (Figure 1-5).

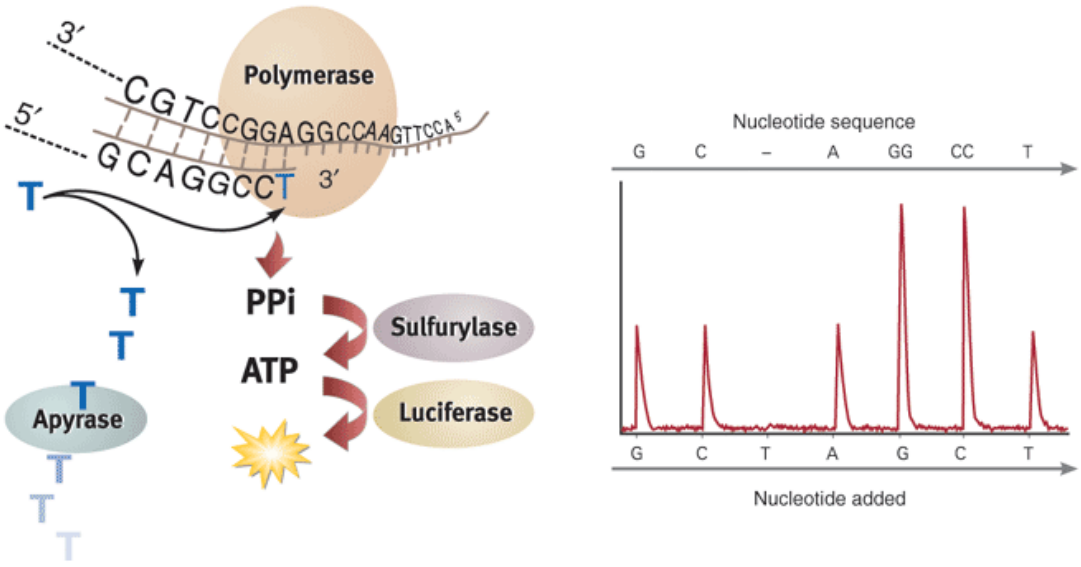


Figure 1-5: The pyrosequencing reaction mechanism (left) and output Pyrogram (right) showing the confirmed DNA sequence. Image courtesy of: England and Pettersson, *Nature Methods* **2** (2005).

DNA methylation analysis by pyrosequencing (BS-pyroseq) has been extensively used for targeted DNA methylation analysis of candidate genomic regions (Baba et al., 2010; White et al., 2006; Xinarianos et al., 2006), as well as a validation tool for 5mC signal detected using other analysis methods (Choufani et al., 2011; Feber et al., 2011; Mirmohammadsadegh et al., 2006; Schatz et al., 2004).

1.4 5hmC profiling methods

1.4.1 Global 5hmC detection

Studies concerning the newly identified oxidative derivatives of 5mC have been hampered by the lack of appropriate methods to identify and profile these modifications; the first 5hmC detection methods allowed only quantification of global 5hmC levels in genomic DNA. Thin-layer chromatography (TLC) assays were first used to confirm the presence of 5hmC in the DNA of mouse Purkinje neurons and granule cells at levels of 0.59% and 0.23% of the total dG content, respectively, whilst 5hmC levels in ESCs were measured at 0.032% of all bases, slightly lower than in the neuronal cells (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Restriction enzyme-based methods have been used to quantify 5hmC in various mouse and human tissues, although these are limited to assaying 5hmC in specific recognition sequences (Kinney et al., 2011; Szwagierczak et al., 2010).

Several methods have employed anti-5hmC antibodies to measure global 5hmC, including immunostaining, dot-blots and enzyme-linked immunosorbent assays (ELISA). Immunostaining of fixed tissue samples (Hackett et al., 2012; Haffner et al., 2011; Jin et al., 2011b; Lian et al., 2012; Ruzov et al., 2011; Szulwach et al., 2011a) and immuno dot-blots using anti-5hmC antibodies (Jin et al., 2011a; Nestor et al., 2012) are perhaps the most widely used methods for quick and easy assessment of the relative distribution of 5hmC across a wide range of cell and tissue types. However, the most sensitive and accurate 5hmC quantification method is undoubtedly liquid chromatography-mass spectrometry (LC-MS), which has been used to measure 5hmC in various mouse tissues and cell lines (Globisch et al., 2010a; Jin et al., 2011b; Kraus et al., 2012; Munzel et al., 2010).

1.4.2 Locus-specific detection

The first developments in locus-specific 5hmC detection were centred around affinity-based DNA enrichment methods followed by PCR, microarray or sequencing analysis (Matarese et al., 2011). The first genome-wide profile of 5hmC used selective chemical labelling of 5hmC coupled to high-throughput sequencing to map 5hmC in the cerebellum of postnatal day 7 and 2.5-month old mice (Song et al., 2011b; Szulwach et al., 2011a, 2011b). This method uses a ß-glucotransferase (ß-GT) enzyme from T4 bacteriophages to label 5hmC with a customised glucose moiety (UDP-6-N_3-glucose), resulting in ß-glucosyl-5-hydroxymethylcytosine (5gmC). Click chemistry is then used for biotinylation of the azide group of 5gmC, allowing high-affinity purification of the biotin-labelled DNA fragments using avidin. Other methods using selective glusosylation and biotin-labelling have been developed which do not require the engineering of a modified glucose moiety. The GLIB (glusosylation, periodate oxidation and biotinylation) method uses ß-GT to glucosylate 5hmC with UDP-glucose, followed by oxidation by sodium periodate to an aldehyde derivative, which can be biotinylated using aldehyde-reactive probes for selective pull-down of 5hmC-containing DNA for sequencing analysis (Pastor et al., 2011). Similarly, glycosylation of genomic DNA can be combined with restriction-enzyme digestion to detect both 5mC and 5hmC. The MspI restriction endonuclease is methylation-insensitive and cleaves CCGG sites containing C, 5mC or 5hmC at the internal cytosine position, but glucosylation of 5hmC protects this site from MspI

digestion (Huang et al., 1982; Josse and Kornberg, 1962; Waalwijk and Flavell, 1978). Therefore comparison of MspI digestion patterns between glycosylated and untreated DNA can be used to quantify the 5hmC content of DNA at CCGG loci. Digestion with HpaII, an isoschizomer of MspI that cannot cleave CCCGG sites when 5mC, 5hmC or 5gmC is present at the internal cytosine position, can provide a measure of the total methylated content of DNA. As a consequence, digestion of glucosylated and untreated DNA in parallel with either MspI or HpaII may be used to quantify 5mC levels by subtraction of the MspI-generated 5hmC estimate from the HpaII estimate of total methylation. Digested DNA may then be analysed using gel electrophoresis or PCR amplification using primers flanking a CCGG recognition site of interest to assess global levels of 5hmC and 5mC (Davis and Vaisvila, 2011; Kinney et al., 2011). Alternatively, digested DNA may be hybridised to a tiling microarray for genome-wide, single-base resolution of modified bases at CCGG sites (Khare et al., 2012).

Antibody-based immunoprecipitation of hydroxymethylated DNA fragments (hMeDIP) was also quickly developed as an alternative enrichment strategy for 5hmC profiling, based on well-established MeDIP protocols for 5mC profiling. The first hMeDIP-seq profiles were generated for wild-type (WT) and mutant mESCs and differentiating embryoid bodies (EBs) using antibodies raised against 5hmC (Ficz et al., 2011; Williams et al., 2011). 5hmC has since been mapped in a variety of cell lines and tissue types using hMeDIP-seq, including mouse PGCs, mouse NPCs, human ESCs, *in vitro* derived human NSCs and dopamine neurons, and human melanoma samples (Hackett et al., 2012; Kim et al., 2014; Lian et al., 2012; Stroud et al., 2011; Tan et al., 2013). 5hmC-containing DNA fragments captured using hMeDIP assays have also been interrogated through hybridization to tiling microarrays for analysis of specific genomic features. For example, 5hmC was mapped in human frontal lobe DNA using the NimbleGen 720K microarray, which covers just over 720,000 CpG sites within CpG islands and promoter regions (Jin et al., 2011a), the NimbleGen 72K ENCODE microarray was used to determine 5hmC enrichment across over 72,000 biologically significant genomic regions, as defined by the ENCODE project, in several normal human tissues, including brain, colon and liver (Nestor et al., 2012), and 5hmC was mapped in WT and Tet1-knockdown

mESCs using NimbleGen whole-genome tiling arrays (Wu et al., 2011). A related IP-based method uses an antibody raised against 5-methylenesulphonate (CMS), the product of BS conversion of 5hmC, to pull down CMS-containing DNA fragments after BS conversion for subsequent analysis using sequencing, PCR or microarrays (Huang et al., 2012; Pastor et al., 2011).

1.4.3 Single-base resolution detection

After publication of several genome-wide 5hmC profiles using a variety of affinity-purification methods, a modification of the traditional bisulphite conversion chemistry made it possible to profile 5hmC at single-base resolution in DNA, resulting in the first quantitative mapping of 5hmC in mESCs (Booth et al., 2012). This selective chemical conversion method, termed oxBS chemistry, includes a DNA oxidation step using potassium perruthenate prior to bisulphite conversion. Oxidation results in the conversion of all 5hmC modifications to 5fC, whilst leaving 5mC unchanged. 5fC behaves as unmodified C during subsequent sodium bisulphite treatment and is fully converted into uracil. Therefore, under oxBS conditions, only 5mC bases are detected as cytosine after sequencing. Subtraction of the oxBS-generated cytosine readout from BS-only datasets allows detection of 5hmC sites. An alternative method for single-base resolution of 5hmC has also been developed, termed TET-assisted bisulphite sequencing (TAB-seq), which also uses a selective oxidation step prior to BS conversion in order to discriminate 5mC from 5hmC (Yu et al., 2012). DNA is first treated with ß-GT to glucosylate 5hmC residues, before using an excess recombinant TET1 protein to fully oxidise the remaining 5mC bases to 5caC, whilst 5gmC is protected from further oxidation. BS treatment leads to the conversion of all 5caC (originally 5mC) and unmodified C to uracil, whilst 5gmC remains protected. After sequencing, only 5gmC positions (originally 5hmC) bases will be read as C, whilst all original 5mC and C positions will be read as thymine (T), allowing quantitative base-resolution analysis of 5hmC in one sequencing reaction. TAB-seq has since been used to profile 5hmC in both human and mouse foetal and adult frontal cortex (Lister et al., 2013; Wen et al., 2014). In addition, single molecule real-time (SMRT) sequencing technology can differentiate between C, 5mC and 5hmC in DNA and has been applied to genome-wide hydroxymethylome profiling in an mESC sample (Flusberg et al., 2010; Song et al.,

2011a). SMRT sequencing is a sequencing-by-synthesis method that monitors the continuous incorporation of phospholabelled nucleotides in newly synthesised DNA molecules through the detection of fluorescent pulse emissions. Measurement of the interpulse duration allows differentiation between modified cytosine bases, although 5hmC labelling with ß-GT and biotin followed by selective enrichment of 5hmC-containing DNA prior to SMRT sequencing greatly increases the detection confidence of this method and reduces the sequencing cost per sample.

1.5 DNA methylation in cell-fate specification

DNA methylation is thought to be essential for maintaining the phenotypic plasticity of a cell in response to both environmental and developmental cues (Feinberg, 2007). The transition from a pluripotent state to a terminally differentiated cell type coincides with many tissue-specific alterations in DNA methylation that have been linked to the regulation of gene expression and alternative transcription (Irizarry et al., 2009; Shiota et al., 2002). Greater differences in DNA methylation levels have been found between tissue types than between individuals; for example, 7.1% of all CpG sites show differential 5mC levels between dermal fibroblasts and CD4+ lymphocytes, compared to differences of only 0.275% and 0.1% between individuals of different age and gender, respectively (Eckhardt et al., 2006). These differentially methylated regions fall within both intragenic and intergenic regions and are preferentially located at highly evolutionarily conserved sequences, suggesting a potential regulatory function.

DNA methylation at CpG-rich promoter regions varies considerably depending on cell and tissue type, suggesting a role in regulating these changing transcriptional profiles during development. Gains in promoter-associated 5mC during development are usually associated with silencing of the corresponding gene, as shown by studies of X chromosome inactivation (Mohandas et al., 1981). An *in vitro* model of neurogenesis demonstrated that mESCs gain CpG-rich promoter methylation during lineage commitment as they differentiate into radial-glial progenitor cells and then into postmitotic neurons. This *de novo* methylation occurred in approximately 2-3% of all interrogated promoters (around 350 out of 15,000), mostly during the transition from mESCs to neuronal progenitor cells, with few further 5mC gains observed in

the differentiation to neurons. In contrast, loss of methylation is seen only at 0.1% (or 22) of the promoters tested. Promoter-associated 5mC gains were also enriched at pluripotency-specific genes, including *Oct4*, *Nanog*, *Sox2* and *Zfp42*, suggesting methylation functions primarily as a repressive mark to suppress the pluripotent potential of cell during development (Mohn et al., 2008).

An array-based analysis of over 14,000 CGIs in human brain, blood, spleen and muscle tissue revealed that less than 10% of all CGIs within promoters are methylated in somatic tissues, but around half of these CGIs show extensive tissue-specific differential methylation patterns and many are associated with genes essential in development, such as homeobox genes (Illingworth et al., 2008). In addition, a MeDIP-based analysis of 5mC in four different mouse tissues (brain, heart, liver and testis) found that most detected regions of 5mC within promoters and CGIs were also tissue-specific differentially methylated regions (T-DMRs), whereas around 10-fold fewer loci were uniformly methylated in all four tissues (460 compared to 4,686 T-DMRs), supporting a functional role for 5mC in regulating tissue specification development (Liang et al., 2011). The application of genome-wide RRBS-seq in a variety of pluripotent and terminally differentiated cells allowed 5mC profiling at around 5% of all CpG sites (approximately 1 million CpGs) in both CpG-rich and CpG-poor regions of the genome. Approximately 8% of all CpG sites became hypermethylated and 2% lost methylation upon ESC differentiation; most of these differentially methylated CpG sites associated occurred at putative distal regulatory elements up to 100kb away from annotated promoters (Meissner et al., 2008a). More recently, this observation has been supported by single-base resolution microarray analysis of 5mC at over 4.5 million CpG sites in a variety of human and mouse tissues, which revealed that the greatest tissue-specific methylation variation is located mostly outside of promoters, in CpG island shores that are located within 2kb of CGIs. Over half of CGI shores are in intergenic regions, over 2kb away from the nearest annotated gene (Irizarry et al., 2009).

Changes in the non-CpG content of cells have also been implicated in differentiation processes; nearly 25% of all methylated cytosines in H1 ESCs are found in non-CpG context, almost all of which seem to be lost upon differentiation via six days of

BMP4 exposure (Lister et al., 2009). This observation indicates such global DNA methylation changes are necessary for the switching of cell fates; in this instance, non-CpG methylation may be a distinguishing feature of stem cells and a significant contributor towards establishing pluripotency.

1.5.1 The potential role of 5hmC in neural development

The discovery of 5hmC as a stable modification in mammalian DNA has led to the hypothesis that this modified cytosine may exert its own gene regulatory function within the genome in an analogous manner to 5mC. It appears there is a correlation between global 5hmC levels and differentiation potential: the 5hmC content of DNA increases as a cell loses its proliferative capacity and transitions from a pluripotent, stem cell-like state to a fully differentiated cell type (Haffner et al., 2011; Nestor et al., 2012; Orr et al., 2012). Elevated levels of 5hmC have consistently been found in tissues of the CNS, where the 5hmC content of DNA is around ten-fold higher than in other somatic tissues or proliferating cell lines (Globisch et al., 2010b; Kriaucionis and Heintz, 2009; Song et al., 2011b). In addition, global 5hmC levels in the mouse cerebellum increase during aging, rising from approximately 0.1% of total nucleotides at postnatal day 7, when extensive cell proliferation is occurring in developing brain, to 0.4% at 2.5 months of age (Song et al., 2011b), with a similar increase of approximately 4-fold being detected between postnatal day 7 and 6-week-old mouse cerebellum (Szulwach et al., 2011a). Unpublished data from Song et al. (2011) also suggests 5hmC levels increase during the *in vitro* differentiation of adult mNSCs.

It has previously been shown that active DNA demethylation is integral to neuronal function and plays a role in learning and memory formation. Locus-specific active DNA demethylation of the brain-derived neurotrophic factor (*BDNF*) promoter occurs in response to neuronal stimulation, releasing promoter-bound MECP2 and resulting in upregulation of *BDNF,* a key regulator of synaptic plasticity (Martinowich et al., 2003; Yoshii and Constantine-Paton, 2010). MECP2 has also been demonstrated to bind 5hmC and 5mC with similar affinities and is the principal 5hmC-binding protein in mouse brain (Mellen et al., 2012), which supports the previous findings of both MECP2 and 5hmC being preferentially located in regions

of highly accessible chromatin and elevated gene expression (Song et al., 2011b; Thambirajah et al., 2012). 5hmC analysis in mouse models of Rett syndrome, a neurological disorder characterized by *MECP2* mutations, found that 5hmC levels in the cerebellum are inversely correlated with MECP2 dosage and that MECP2 binding to 5mC can inhibit TET1-mediated oxidation of 5mC (Szulwach et al., 2011a). Furthermore, combined MeDIP- and hMeDIP-seq analysis of the epigenetic changes occurring within the anterior *HOXA* cluster (*HOXA1-6*) during the neural differentiation of a human embryonal carcinoma cell line showed that 5mC to 5hmC conversion was a critical factor in the activation of this gene set (Bocker et al., 2012). Taken together, these observations all suggest a unique role for 5hmC in transcriptional regulation during neurodevelopment and brain maturation.

1.5.2 Development of the central nervous system

Neurons and macroglia are the two main cell types present in the brain and spinal cord, which together make up the mammalian CNS. These organs originate from a multipotent, self-renewing NSC population present in the neural tube of the developing embryo, which is comprised of a single layer of neuroepithelial cells showing strong apical-basal polarity (Götz and Huttner, 2005). CNS development occurs under strict spatiotemporal control, with embryonic neurogenesis taking place before the onset of gliogenesis. Initially, neuroepithelial cells undergo an expansion phase of symmetric cell divisions, before switching to asymmetric divisions which each produce a neuronal cell and a daughter stem cell (Figure 1-6). As neurogenesis progresses, neuroepithelial cells are gradually replaced by radial glial cells and basal progenitor cells, two further NSC subtypes with restricted lineage potential (Temple, 2001).

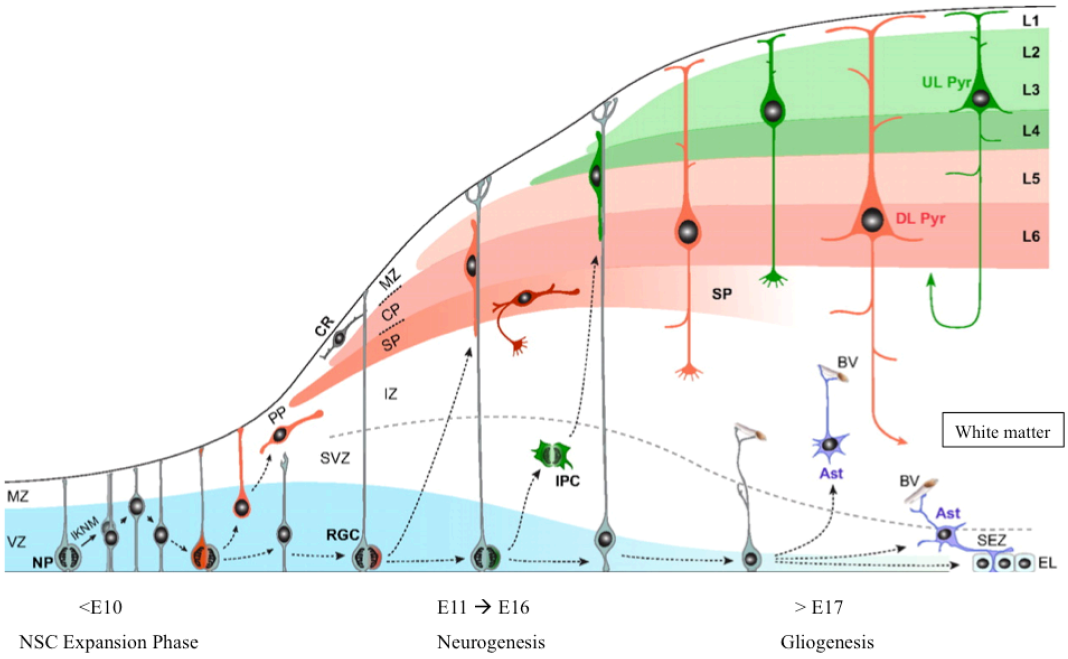


Figure 1-6: Neurogenesis in the developing mouse CNS. Neural development in the mammalian CNS is divided into three major phases that encompasses an NSC expansion phase, a neurogenic phase and a gliogenic phase. Abbreviations: E10-17 = embryonic day 10-17. NP = neural progenitors (neuroepithelial cells). IKNM = interkinetic nuclear migration; MZ =marginal zone; VZ = ventricular zone; SVZ = subventricular zone; SEZ = subependymal zone; PP= preplate; SP =subplate; CP = cortical plate; CR = Cajal-Retzius neuron; L1-L6= neuronal layers; UL Pyr = upper layer pyramidal neurons; DL Pyr = deep layer pyramidal neurons; RGC = radial glial cell; IPC = intermediate progenitor cell; Ast = astrocytes; EL = ependymal cells; BV = blood vessels. Image adapted from: Kwan et al., *Development* **139,** 1535-1546 (2012).

Concomitant with the onset of neuronal differentiation is the transition of the neuroepithelium into a multi-layered tissue. Neurons migrate towards the basal surface, forming a neuronal layer, whilst the most apical cell layer containing radial glial and basal progenitor cells is referred to as the ventricular zone (VZ). Radial glial cells and basal progenitor cells continue to divide asymmetrically to produce both neurons and daughter stem cells, before finally switching to symmetric, neurogenic cell division towards the end of neurogenesis. The switch from neurogenesis to gliogenesis coincides with the development of a subventricular zone (SVZ), a new cell layer basal to the VZ formed by a subset of basal progenitors (Götz and Huttner, 2005; Temple, 2001). Gliogenesis occurs mostly in the SVZ of the developing telencephalon, resulting in the formation of the two main types of macroglial cell: astrocytes and oligodendrocytes. The SVZ remains in the lateral ventricle of the adult mammalian brain and is a source of adult NSCs (Lois and Alvarez-Buylla, 1993; Reynolds and Weiss, 1992).

1.5.3 Neural development in the adult mammalian brain

A subset of cells in the mammalian CNS have been found to retain NSC-like properties throughout adulthood and can differentiate to produce both neurons and glia (Lois and Alvarez-Buylla, 1993; Morshead et al., 1994; Reynolds and Weiss, 1992; Weiss et al., 1996). They can be found in either the SVZ of the lateral ventricle (Figure 1-7) or in the subgranular zone (SGZ) of the hippocampus. Adult NSCs in the lateral ventricle can differentiate to produce interneurons, which migrate to the olfactory bulb, whilst NSCs of the hippocampus produce granule cells, which reside in the dentate gyrus (Alvarez-Buylla and Lim, 2004). These cells are known to show the structural and molecular properties of astrocytes and are often described as astroglial or radial glial cells (Doetsch et al., 1999). Adult NSCs have been successfully isolated from the adult mouse brain and established *in vitro,* where they may be induced to differentiate into the three major neural lineages (astrocytes, oligodendrocytes and neurons), providing a useful and convenient source of neural stem cells for research into neurodevelopmental and disease processes (Glaser et al., 2007; Pollard et al., 2006).

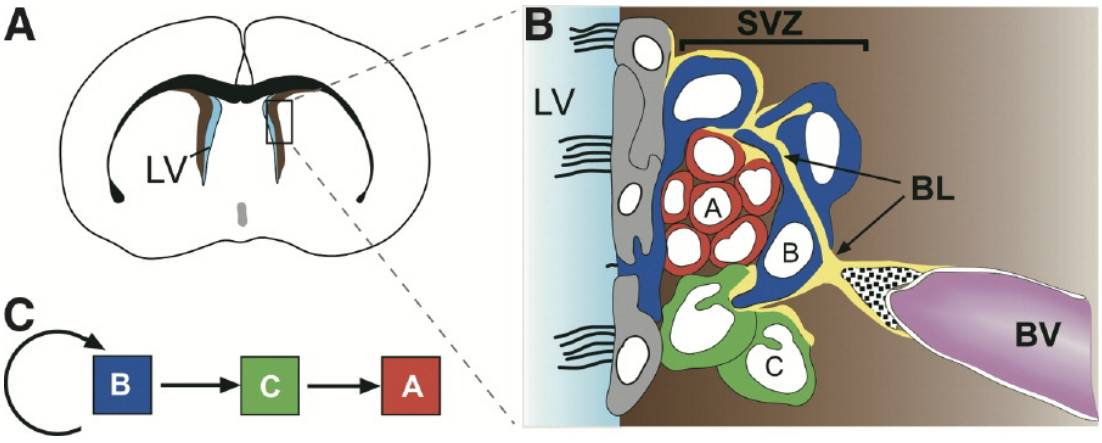


Figure 1-7: Location of NSCs in the adult mouse brain. (A) Cross-section of an adult mouse brain shown the location of the lateral ventricles (LV), where NSCs are found**. (B)** The SVZ layer in the wall of the LV. (C) The interrelationship of the 3 major cell types found in the SVZ. B cells are astrocyte-like cells that behave as NSCs, have self-renewing and differentiation capacity and are frequently GFAP-positive. They can generate C cells, a type of transit-amplifying cell that differentiate into A cells (neuroblasts), which can migrate to the olfactory bulb and function as interneurons. Both C and A cells are GFAP-negative. Image courtesy of: Alvarez-Buylla and Lim, *Neuron* **41(5),** 683-686 (2004).

1.5.4 The role of astrocytes in the central nervous system

Astrocytes are a diverse group of glial cells present throughout the CNS that are divided into two main subgroups, protoplasmic and fibrous astrocytes, based on differences in morphology and location within the brain (Sofroniew and Vinters, 2010). Protoplasmic astrocytes are found in grey matter and are linked to neuronal cell bodies and synapses, whilst fibrous astrocytes are located within white matter and are typically linked to neuronal axons. Traditionally, glial cells have been thought to outnumber neurons by a factor of ten, but more recently a study of cell number in human brain tissue found approximately equal numbers of neurons and glia (Azevedo et al., 2009), although ratios vary dramatically between different brain regions. The cerebral cortex had the highest glial-to-neuron ratio at approximately 4:1, whereas this ratio is reversed in the cerebellum.

The functions of astrocytes are not well understood, but it is generally accepted that astrocytes play a critical function in the regulation of metabolism within the CNS. Astrocytes contain glycogen granules that serve as the main energy stores for neuronal cells and may be transferred to adjacent cells during periods of high synaptic activity or hypoglycaemia (Brown et al., 2004; Pellerin et al., 2007; Voutsinos-Porche et al., 2003). Astrocytic processes form connections to both blood vessels and neurons, allowing the regulation of local blood flow in response to neuronal activity (Iadecola and Nedergaard, 2007; Koehler et al., 2009). There is also evidence for a role in synaptic plasticity and signalling. Astrocytes express cell-surface receptors for many neurotransmitter molecules and may release chemical signals of their own into the extracellular space, including glutamate, d-serine and ATP, which may regulate the function of neighbouring neurons (Haydon and Carmignoto, 2006). Astrocytic processes have also been shown to physically interact with many thousands of neuronal synapses and dendrite branches, providing a mechanism through which they may influence synaptic transmission and regulate the homeostasis of synaptic interstitial fluid (Halassa et al., 2007; Ogata and Kosaka, 2002). Astrocytes are also thought to be important in guiding neuronal axon migration in the developing CNS (Powell and Geller, 1999) and to contribute to normal myelination of axons within white matter (Lutz et al., 2009). The increasingly diverse range of observed astrocytic functions suggests they are not

simply neuronal support cells but instead represent a group of highly specialised neural cells with critical importance in the healthy CNS.

1.6 Aberrant DNA methylation phenotypes and disease

Alongside changes in the distribution of histone modification and chromatin-modifying proteins, aberrant DNA methylation has been linked to a host of developmental abnormalities and disease states (Feinberg, 2007). The importance of normal methylation patterns in development is highlighted by the embryonic lethal phenotype observed in embryos with homozygous mutation of *DNMT1* (Li et al., 1992). The epigenetic deregulation observed in many human disease states is often the result of genetic mutations that result in loss of imprinting or inactivation of a core component of the DNA methylation machinery, such as the DNMT enzymes or MBD proteins. These locus-specific mutations have been linked to many developmental disorders that frequently manifest as neurological disease, including Prader-Willi and Angelman syndrome (Buiting et al., 1995; Horsthemke and Buiting, 2006), Rett syndrome (Guy et al., 2001; Psoni et al., 2012), Huntington's disease (Ng et al., 2013), Alzheimer's disease (Irier and Jin, 2012) and ICF (immunodeficiency, centromere instability and facial anomalies) syndrome (Xu et al., 1999). 5mC deregulation is also commonly observed during tumourigenesis, both in the form of global and site-specific hypo- and hyper-methylation events that are associated with the transcriptional deregulation and epigenetic plasticity observed in cancer cells (Feinberg and Vogelstein, 1983; Jones and Baylin, 2007). Promoter hypermethylation is observed at many tumour suppressor genes that are silenced in cancer, such as the retinoblastoma gene *RB* and the cyclin-dependent kinase inhibitors *CDKN2A* and *CDKN2B* (Greger et al., 1989; Herman et al., 1997), whilst many oncogenes are activated by hypomethylation, including *HRAS* and the cyclin D2 gene, *CCND2* (Feinberg and Tycko, 2004; Feinberg and Vogelstein, 1983; Wilson et al., 2007).

Recently, 5hmC regulation has also been implicated in disease progression, particularly cancer. Global hypo-hydroxymethylation has been observed in many cancer tissues compared to their normal counterparts (Haffner et al., 2011; Jin et al., 2011b; Kraus et al., 2012; Kudo et al., 2012; Lian et al., 2012; Yang et al., 2012) and

has also been associated with poorer prognosis and reduced survival in MDS, glioblastoma and melanoma (Lian et al., 2012; Liu et al., 2013a; Orr et al., 2012). Mutations in either isocitrate dehydrogenase genes 1 and 2 (*IDH1/2*) are often observed in astrocytomas, glioblastomas and acute myeloid leukaemia (AML), leading to TET2 inhibition and a hypermethylation phenotype, with corresponding hypo-hydroxymethylation (Figueroa et al., 2010). Reduced expression of TET1 and TET2 have also been extensively reported in various cancers, including AML, myelodysplastic syndrome (MDS), glioblastoma, prostate and breast cancer, further linking the 5hmC pathway to disease progression (Abdel-Wahab et al., 2009; Delhommeau et al., 2009; Hsu et al., 2012; Langemeijer et al., 2009; Muller et al., 2012).

1.7 DNA methylation as a therapeutic target

Given the widespread epigenetic deregulation seen in many disease phenotypes, the restoration of normal DNA methylation patterning is a promising avenue of research in the development of multiple therapies. So far, the U.S. Food and Drug Administration (FDA) have approved two DNA demethylating drugs for clinical use in the treatment of MDS and AML (Kaminskas et al., 2005; Kantarjian et al., 2006). 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine (DAC), also known by their trade names Vidaza and Dacogen, respectively, are cytosine analogues that incorporate into newly synthesised DNA molecules, where they act as DNMT inhibitors through irreversible covalent binding of the enzyme, thereby depleting the cell of active DNMT (Figure 1-8). However, it is unknown whether their therapeutic usefulness stems from an ability to truly reprogram cancerous cells or if these DNA demethylating agents simply trigger cell death due to cytotoxicity (Kaiser, 2010).

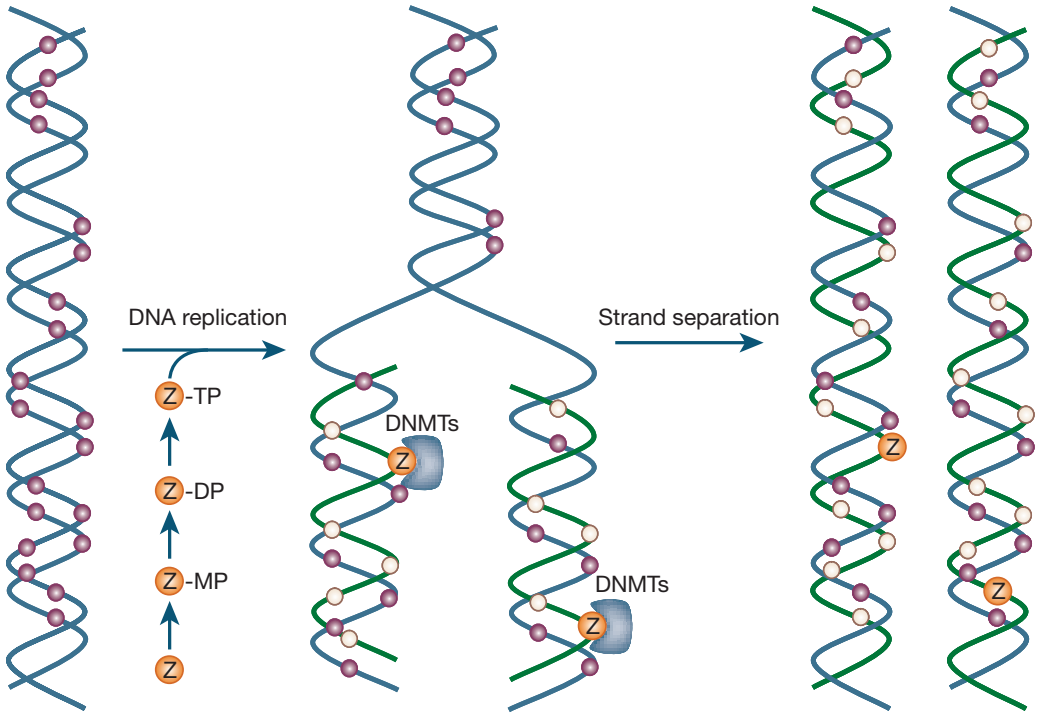


Figure 1-8: Mechanism of action of 5'-azacytidine and 5-aza-2'-deoxycytidine as a DNA demethylating agent. The molecule labelled "Z" represents the cytosine analogues 5-aza and DAC. The cytosine analogues incorporate into newly synthesised daughter strands, where they may covalently bind DNMT molecules, sequestering their enzymatic activity. The DNA demethylation activity of these epigenetic drugs depends on DNA replication, allowing selective targeting of proliferating cells. Image courtesy of Egger et al., *Nature* **429,** 457-463 (2004).

Whilst shown to be effective in treating certain cancers, these DNA demethylation drugs are thought to be non-specific in both their cellular and genomic targets and there is concern that they may induce hypomethylation at normally methylated sites, potentially activating oncogenes in normal tissue, as seen in mouse models of induced hypomethylation (Gaudet et al., 2003). However evidence is emerging that these DNMT inhibitors act in a non-random and reproducible manner, preferentially targeting CpG dinucleotides outside of CGIs, whilst other genomic regions are consistently resistant to DNA demethylation (Hagemann et al., 2011). The potential of developing targeted DNA demethylating agents is explored in Chapter 5. Interestingly, treatment with cytosine analogues may also affect 5hmC levels, as noted in a recent study using DAC treatment in a leukaemia cell line, which observed both increased global 5hmC and increased TET1 binding in DNA (Chowdhury et al., 2015). As reduced 5hmC levels have emerged as a frequent hallmark of cancer, this study provides an alternative mechanism through which DNA demethylating agents may be exerting their therapeutic capacity and warrants further investigation into the

possibility of using 5hmC and the TET enzyme family as novel targets in the development of epigenetic therapies.

1.8 Project aims

1) Investigation into 5mC and 5hmC dynamics during neural differentiation using genome-wide, MeDIP- and hMeDIP-seq profiling methods to analyse the changing distribution of both modifications and explore the potential function of 5hmC in the mammalian genome.

2) Development of a novel method for 5hmC detection at single-base resolution through adaptation of oxidative bisulphite chemistry for use in combination with the Infinium HumanMethylation450 BeadChip.

3) Assessment of the potential use of photocaged DNMT inhibitors for targeted DNA demethylation treatment using an *in vitro* cell culture assay combined with genome-wide 5mC analysis.

Chapter 2 Materials and methods

2.1 Cell lines

The JM8 cell line (WT agouti C57BL/6N mouse ESCs) was obtained from the laboratory of Dr William Skarnes (Sanger Institute, Hinxton, Cambridge) and a *Sox1-GFP* knock-in E14Tg2a (46C) mouse ESC line was obtained from the laboratory of Dr Paul Bertone (EMBL-EBI, Hinxton, Cambridge). Both were cultured according to the protocol described in section 2.3. Primary mouse NSCs were previously isolated from the subventricular zone of an adult mouse brain (C57BL/6N strain) and established in culture by Dr Stefano Bartesaghi at the Samantha Dickson Brain Cancer Unit of Professor Paolo Salomoni (UCL Cancer Institute, University College London). These cells were cultured as described in section 2.3.4. The Saos-2 human osteosarcoma cell line was obtained from the laboratory of Dr Sandra Strauss (UCL Cancer Institute, University College London) and cultured as described in section 2.3.6.

2.2 Tissue samples

Two 6-month old, WT C57BL/6N female mice were sacrificed using CO_2 euthanasia for the dissection of brain tissue samples. Tissue samples were taken from the cortex, cerebellum and olfactory bulb of both mice to serve as biological replicates.

All human tissue samples were handled in accordance with the ethical standards of the Helsinki declaration (1964, 2008) of the World Medical Association. Human brain genomic DNA from a 33 year-old male was obtained from AMSBIO (#D1234035, lot #B701205; referred to as 'brain1'). Human frontal cortex DNA was collected from a 54-year-old female according to ethical approval obtained from the Berkshire Local Research Ethics Committee (REC 08/H0505/165), with informed written consent (referred to as 'brain2'). Pooled (male and female) whole blood DNA was obtained from blood donations of volunteers from the UCL Cancer Institute.

2.3 Cell culture

2.3.1 Cell counting

In order to calculate accurate cell seeding densities when passaging, cell numbers were quantified using a Vi-CELL XR (Beckman Coulter) automated cell viability analyser, which uses the trypan blue exclusion method to count viable cells within a 500µl volume.

2.3.2 Embryonic stem cell culture

Preparation of plastic culture dishes: 2% gelatin solution (Sigma-Aldrich; #G1393-100ML) was diluted to 0.1% with 1X Dulbecco's phosphate-buffered saline (DPBS; Gibco; #14190-169) and stored at 4°C. All plasticware used for mESC culture was first coated with a 0.1% gelatin solution for 30 minutes at room temperature. Excess gelatin solution was aspirated from the surface and let to air dry for approximately 1 minute before cell seeding.

Cell expansion: JM8 and 46C cells were initially cultured as an adherent monolayer in LIF- and serum-containing growth medium, comprising of knockout DMEM (Gibco; #10829-018), 2mM L-glutamine (Gibco; #25030-081), 100µM ß-mercaptoethanol (Gibco; #21985-023), 10% foetal bovine serum (FBS) (Gibco; 10500-064) and 10^3 U/ml LIF (Millipore; #ESG1106). After a couple of passages, both mESC cell lines were switched to serum- and LIF-free "3i" medium using iSTEM (StemCells, Inc.; #SCS-SF-ES-01) and subsequently expanded solely under these conditions. All cells were incubated at 37°C in 5% CO_2.

Cell passaging: 30 minutes prior to passaging, aliquots of Accutase (Sigma-Aldrich; #A6964) and growth medium were pre-warmed in a water bath at 37°C and 1X DPBS solution equilibrated to room temperature. New plasticware was coated with 0.1% gelatin solution and left at room temperature for a minimum of 30 minutes under a culture hood until needed. To passage mESCs, all growth media was aspirated from the culture flask and retained for use as 'wash media' in subsequent steps. The flask's surface was washed gently with 1X DPBS, taking care not to disrupt the cell layer. The remaining DPBS was aspirated away and pre-warmed Accutase added to the flask (approximately 0.8ml for a T25 flask, or 2.5ml for a T75

flask). The flask was tilted gently to ensure the entire cell surface was covered with Accutase solution, before returning the flask to an incubator (37°C, 5% CO_2) for a maximum of 5 minutes. Cells were observed under a microscope to check for detachment and then a small amount of 'wash media' (2ml for a T25 flask, 5-10ml for a T75 flask) was added back to the flask to resuspend the detached cells. The cell suspension was then transferred to a new 15ml falcon tube and spun down for 3 minutes at 1200rpm to pellet the cells. The remaining supernatant was aspirated away and the cell pellet dissociated and resuspended in fresh growth media. The cells were then split 1:6 in freshly-coated culture flasks or plates, as needed. Finally, all flasks were topped up with pre-warmed growth media to a total volume of 5ml for T25 flasks, or 15ml for T75 flasks.

Cell freezing: To freeze down cell aliquots, freezing medium was prepared by adding 10% DMSO to growth medium (at room temperature) immediately prior to use. Cells were dissociated after reaching approximately 70-80% confluence and spun down to a pellet as described above, before being resuspended in freezing medium in the appropriate volume:

Plasticware	Freezing media volume	Number of cryovials needed
6-well plate	1ml (per well)	2
25cm^2 flask	2.5ml	3
75cm^2 flask	7.5ml	9

0.5ml of the resulting cell suspension was transferred into each cryovial and immediately placed on dry ice for 10 minutes, before being transferred to a -80°C freezer overnight. The next day, cryovials were transferred to liquid nitrogen for long-term storage.

2.3.3 Neural differentiation

Generation of a stable, self-renewing mNSC line: Neural differentiation of mESCs was performed using the *Sox1-GFP* knock-in E14Tg2a (46C) mESC line in order to monitor the appearance of SOX1-positive neural precursor cells. 46C mESCs growing under 3i conditions were dissociated, spun down and resuspended in RHB-A media (without growth factors or serum) and replated at 1 x 10^4 cells/cm^2 in

gelatin-coated T25 flasks. Media was exchanged every other day and cells were monitored daily for the appearance of GFP (indicating *Sox1* expression) using a fluorescence microscope, which typically occurred around day 7-8 of differentiation and peaked at day 10, when neural rosette structures were clearly visible throughout the culture. At this stage, cells were dissociated with Accutase, spun down and resuspended in freshly prepared RHB-A media with 10ng/ml EGF and FGF-2. Cells were then replated on laminin-coated plasticware at the same cell density and expanded as embryonic mNSCs, following the protocol described in section 2.3.4. Growth factor concentrations were increased to 20ng/ml for one or two passages if cells exhibited low viability upon this replating step.

Continuous neuronal differentiation: three variations on a neuronal differentiation protocol were tested to produce terminally differentiated neurons from mESC cultures, referred to as methods A, B and C (see Chapter 3). In method A, mESCs grown under 3i conditions were replated in RHB-A media (without supplements) at 0.5-1.5 x 10^4 cells/cm^2 on gelatin-coated plasticware to initiate differentiation. Cells were maintained under these differentiation conditions for 3 weeks, when no further changes in cell morphology or cell number were observed. In method B, mESCs were replated in RHB-A (without supplements) at 0.5-1.5 x 10^4 cells/cm^2 on gelatin-coated plasticware to initiate differentiation and monitored for *Sox1-GFP* expression, upon which they were dissociated, resuspended in fresh RHB-A and replated on a laminin-coated substrate at 4-5 x 10^4 cells/cm^2 and monitored for the appearance of neuronal morphology, such as axonal projections and dendritic branches. In method C, stably renewing embryonic mNSCs are derived as described above and cultured in RHB-A with FGF-2 and EGF, before being replated at approximately 2-3 x 10^4 cells/cm^2 (approximately 6.5 x 10^5 cells in a T25 flask) on a laminin substrate and withdrawing EGF from the expansion media. Cells are maintained in these conditions for 7 days, before withdrawing FGF-2 from the media and continuing the culture in RHB-A only for a further 7 days (minimum) and monitored for the appearance of neuronal-like cells.

2.3.4 Neural stem cell culture

Preparation of cell culture reagents: A 1:100 dilution of laminin (1mg/ml, Sigma-Aldrich) was prepared in sterile-filtered 1X DPBS and 5ml aliquots of the resulting 1µg/ml solution stored at -20ºC. Recombinant human FGF-basic (FGF-2, PeproTech; #100-18B) was reconstituted in 10µl 5mM Tris (pH 7-7.6) and diluted 1:100 in a 0.1% solution of bovine serum albumin (BSA) made up in 1X DPBS. The resulting10µg/ml FGF-2 solution was stored in aliquots at -20ºC and used at a final concentration of 10ng/ml. Recombinant human EGF (EGF, PeproTech; #AF-100-15) was reconstituted in 100µl sterile H_2O and a 1:100 dilution prepared in 0.1% BSA solution (diluted in 1X DPBS). The resulting 10µg/ml EGF solution was stored in aliquots at -20ºC and used at a final concentration of 10ng/ml.

Preparation of plastic culture dishes: All new flasks and tissue culture plates were coated with a 1µg/ml laminin solution (approximately 1ml added to a T25 flask) for a minimum of three hours at room temperature prior to cell seeding. All plasticware was placed on a bench-top orbital shaker for the incubation period to ensure all surfaces were evenly coated with laminin.

Cell expansion: Primary and embryonic mNSCs were grown as adherent monolayers in laminin-coated flasks. Expansion medium was comprised of RHB-A (StemCells, Inc.; SCS-SF-NB-01), a serum-free chemically defined neural stem cell medium, along with epidermal growth factor (EGF, PeproTech; #AF-100-15) and fibroblast growth factor (FGF-2, PeproTech; #100-18B), both at a final concentration of 10ng/ml. All cells were incubated at 37°C in 5% CO_2.

Cell passaging: mNSCs were passaged after reaching 75-90% confluence and typically split 1:5 into freshly-coated laminin flasks using the same procedure as previously described for mESCs in 2.3.2.

Cell freezing: Both primary and embryonic mNSCs were frozen down as described for mESCs in section 2.3.2, using freezing medium comprised of RHB-A, 10ng/ml FGF-2, 10ng/ml EGF and 10% DMSO.

2.3.5 Astrocytic differentiation

Astrocytic differentiation of stably self-renewing embryonic or adult mNSCs was triggered through the simultaneous withdrawal of EGF and FGF-2 and addition of 10% FBS (Gibco; #10500-064) to RHB-A media. Aliquots of 10% FBS-containing differentiation media were sterile-filtered using a 0.2µm filter prior to use. After aspiration of the mNSC expansion media, mNSC cultures were washed gently with 1X DPBS prior to addition of the differentiation media, taking care not to let cells dry out. Differentiation was initiated when mNSCs cells were at 50-75% confluence within a tissue culture plate flask or plate. Differentiation media was exchanged every third day for the duration of the differentiation period and cells were collected for DNA extraction (or fixed in paraformaldehyde for immunostaining) 14 days after initiating differentiation.

2.3.6 Saos-2 cell culture

The Saos-2 cell line was grown as an adherent monolayer in RPMI-1640 medium (Lonza; #12-702F) supplemented with 10% FBS (Gibco; #10500-064) on uncoated tissue culture plasticware. Cells were passaged as described above for mESCs (see section 2.3.2), although Trypsin (Gibco; #25300-062) was used in the place of Accutase to dissociate cells during passaging.

2.3.7 Saos-2 cell line treatments with cytosine analogues

Cytosine analogues: 5-aza and DAC compounds were purchased from Sigma-Aldrich (#A2385 and #A3656, respectively). Caged-DAC and caged-aza compounds were synthesised by Mikiembo Kukwikila and Ha Phuong Nguyen (Department of Chemistry, University College London) and provided as 10mM solutions in acetonitrile. Unless otherwise stated, all compounds were stored as 1mM solutions at -20°C and diluted to 100µM in acetonitrile just prior to use.

In vitro cell treatments: Saos-2 cells were seeded into 24-well tissue culture plates at an initial density of approximately 10,000 cells per well 24 hours before the first treatment dose. 500µl of culture media was added per well and exchanged every 24 hours for fresh media, prior to addition of a new treatment dose. 100µM solutions of 5'-azacytidine (5-aza), 5'-aza-2'-deoxycytidine (DAC) caged-aza and caged-DAC

compounds were added directly to fresh growth media in the appropriate wells every 24 hours for the duration of the treatment period. Irradiation of uncovered tissue culture plates was carried out at 365nm for one hour immediately after the required compound dosage was added to the media. A 65µW/cm^2 UV lamp (Mineralight; #UVGL-58) was used to irradiate uncovered 24-well culture plates at room temperature under a culture hood, within a UV black box. Control plates that were not receiving irradiation treatment were incubated at room temperature under the culture hood for the duration of the 1-hour treatment. Care was taken to minimise light exposure to these control cells by covering the plates in foil.

2.3.8 Live cell imaging

A live cell imaging system was used to monitor the 5-aza, DAC, caged-aza and caged-DAC treatments to Saos-2 cells (see Chapter 5, section 5.3). Cells were monitored using an IncuCyte (Essen BioScience), which allowed multiple cell confluence measurements and phase contrast images to be taken in each well of multi-well culture plate once every two hours.

2.4 EdU assay

EdU incorporation assays were performed using the Click-iT EdU Imaging Kit (Molecular Probes; #C10086) with Alexa Fluor 488, 594 and 647 azides according to the manufacturer's instructions, with a few minor modifications. For analysis of mESCs and mNSCs, cells were seeded into 4-well plates under expansion conditions (see section 2.3.2 and 2.3.4, respectively) 24 hours prior to EdU labelling. For analysis of astrocytes, 40-60,000 mNSCs were seeded per well of laminin-coated 4-well plates, incubated in expansion media for 24 hours and subsequently switched to differentiation media for 14 days (see section 2.3.5) prior to addition of EdU. A 20µM EdU solution was freshly prepared through a 1:500 dilution of 10mM EdU in pre-warmed culture media. Cells were incubated with 20µM EdU at 37°C for 2 hours before fixation at room temperature for 10 minutes in 4% paraformaldehyde (PFA), prepared using a 1:3 mixture of 16% PFA (Alfa Aesar; #43368) with 1X DPBS. After aspiration of the PFA solution, cells were gently washed twice with 0.1% PBST, composed of 0.1% Triton X-100 (Sigma-Aldrich) dissolved in 1X DPBS solution. For permeabilisation, approximately 200µl of fresh 0.1% PBST was added

to each well and incubated for 20 minutes at room temperature prior to addition of 200µl per well of the Click-iT reaction cocktail, prepared as described in the manufacturer's protocol. Cells were covered with foil and incubated in the dark for one hour at room temperature and then washed twice with 0.1% PBST. Cells were then subjected to immunocytochemical staining (as described in section 2.5) for the appropriate phenotypic markers (OCT4 for mESCs, nestin for mNSCs and GFAP for astrocytes) prior to visualisation.

2.5 Immunocytochemistry

Cells were plated as an adherent monolayer at the desired density in 4-well tissue culture plates (Thermo Scientific Nunc; #176740) and incubated at 37°C for a minimum of 24 hours prior to fixation. For immunostaining of mESCs and mNSCs, between 50-70,000 cells were typically seeded into each well 24 hours prior to fixation. For immunostaining of astrocytes, mNSCs were seeded at a lower density (between 30-50,000 cells per well) and incubated in expansion media for 24 hours before undergoing differentiation for 14 days prior to fixation (see section 2.3.5 for further details). For immunocytochemical analysis, growth media was aspirated from the surface of the cells and each well washed twice with 1X DPBS, taking care not to disturb the cell layer.

After aspirating away the remaining wash solution, 200µl of 4% PFA (prepared as described in section 2.4) was added to each well and incubated at room temperature for 10 minutes. After aspirating off the remaining fixative, cells were gently washed twice with 1X DPBS before addition of 200µl of freshly prepared 0.1% PBST to each well for permeabilisation. Cells were incubated in 0.1% PBST for a minimum of 5 minutes at room temperature (or stored at 4°C for up to 1 week) before proceeding with immunostaining.

Immunocytochemical staining was performed by aspiration of the 0.1% PBST solution and addition of 200µl blocking solution (0.1% PBST with 4% FBS and 1% BSA) to each well. Fixated cells were incubated with block solution for one hour at room temperature, before replacing block solution with primary antibody dilutions made up using fresh block solution (e.g. 1.5µl of Ab to 1500µl blocking solution for

a 1:1000 dilution). Primary antibodies used for immunocytochemical analysis were a rabbit anti-GFAP IgG antibody (Sigma-Aldrich; #G9269), a mouse anti-nestin IgG_1 antibody (Millipore; #MAB353), both at a 1:800 dilution, and a mouse anti-Oct-3/4 (C-10) IgG_{2b} antibody (Santa Cruz Biotechnology, Inc.; #sc-5279), at a 1:400 dilution. Cells were covered with foil and incubated with the primary antibody solution overnight at 4°C.

After aspiration of the primary antibody solution, cells were gently washed twice with 0.1% PBST, taking care not to let the cells dry out between steps. 200µl of a secondary antibody dilution (in block solution) was then added to each well. Secondary antibodies used for immunocytochemical analysis were an Alexa Fluor 488 donkey anti-rabbit IgG (H+L) antibody (Invitrogen; #A21206) and an Alexa Fluor 555 goat anti-mouse IgG (Invitrogen; #A21422), both at a final dilution of 1:1000. Cells were covered with foil and incubated at room temperature for 45 minutes before aspirating the antibody solution and washing twice with 0.1% PBST. DAPI counterstaining was performed by incubating cells with a 1:10,000 DAPI dilution in 0.1% PBST for approximately 10 seconds before quickly aspirating away all remaining solution. A final wash with 0.1% PBST was performed before covering each well with fresh 0.1% PBST for visualisation or storage at 4°C. Immunostaining was visualised using an Axio Observer Z1 (Carl Zeiss, Ltd.) fluorescence microscope and images taken using AxioVision software.

2.6 DNA extraction

DNA samples for MeDIP- and hMeDIP-seq analysis were prepared from frozen cell pellets (approximately 3 x 10^7 cells) using a DNA/RNA dual extraction protocol (Qiagen AllPrep DNA/RNA Mini Kit; #80204). In order to maximise yield, the extraction protocol was performed on 3 separate aliquots per cell sample due to the large amount of available starting material. First, cell pellets were thawed on ice before being re-suspended in 1800µl RLT plus buffer, pipette-mixed thoroughly and then divided into 3 x 600µl samples for subsequent processing. Each 600µl cell suspension was transferred to an individual QIAshredder spin column (Qiagen; #79654) and spun at maximum speed (13,200rpm, or 16,100 x g) in an Eppendorf

microcentrifuge. The lysate was then used for dual DNA/RNA extraction according to the manufacturer's instructions.

DNA extraction from mouse and human tissues was performed using the DNeasy Blood & Tissue Kit (Qiagen; #69504) according to the manufacturer's recommendations.

2.7 Enzyme-linked immunosorbent assays (ELISA)

A 5mC ELISA assay (5-mC DNA ELISA Kit, Zymo Research; #D5325) was performed according to manufacturer's instructions, with the DNA-coated plate incubated at 4ºC overnight rather than at 37ºC for 1 hour. After addition of HRP developer, the plate was incubated at room temperature for 1 hour prior to absorbance measurements. Final absorbance readings were measured at 450nm on a VarioSkan Flash multimode platereader (Thermo Scientific).

2.8 MeDIP- and hMeDIP-seq

2.8.1 DNA quantification

DNA samples were quantified using a NanoDrop 1000 spectrophotometer and aliquots of approximately 100ng were analysed using agarose gel electrophoresis to check that all samples contained high-molecular weight (>10kb) DNA and were free from RNA contamination. Quantification of low concentration DNA samples was carried out using the Qubit High Sensitivity fluorometric assay kit (Invitrogen; #Q32851).

2.8.2 DNA shearing

DNA samples were sonicated using a Biorupter (Diagenode) in 1.5ml TPX polymethylpentene microtubes (Diagenode; #M-50001). 1.5µg DNA was added to each tube in a total volume of 150µl elution buffer. Each sample was sonicated on the 'high' setting (320W) for a total of four 15-minute cycles (30 seconds on, 30 seconds off) followed by an additional 5-minute cycle. Between cycles, the sonication bath was cooled with ice for 5 minutes to prevent overheating. All sonicated DNA samples were run on an Agilent Bioanalyzer 2100 using a High

Sensitivity DNA chip at a 1:10 dilution to check the majority of fragments fell within the desired range of 150-350bp.

2.8.3 Library preparation (part 1)

MeDIP and hMeDIP DNA libraries were prepared using the NEBNext DNA Library Prep Master Mix Set for Ilumina (New England Biolabs; #E6040L) and the Multiplexing Sample Preparation Oligonucleotide Kit (Illumina; #PE-400-1001). Sonicated DNA underwent DNA end repair, dA-tailing and adapter ligation reactions according to the manufacturer's protocol and using Illumina Index paired-end (PE) adapters. DNA purification steps were performed using Agencourt AMPure XP (Beckman Coulter; #A63881) beads and the final elution step was performed in a final volume of 40µl buffer EB (Qiagen; #19086).

2.8.4 DNA immunoprecipitation

The DNA immunoprecipitation steps were automated on an SX-8G IP-Star using the Auto MeDIP kit (Diagenode; #AF-Auto01-0100) and Auto hMeDIP kit (Diagenode; #AF-Auto02-0016), according to the manufacturer's instructions, with one minor modification for the auto MeDIP reactions: two lambda DNA fragments, one unmethylated ("20CpG") and one fully methylated ("15CpG"), were added to each sample at a concentration of 0.2pM to act as spiked-in control DNA. For auto hMeDIP reactions, the spike-in DNA controls provided by the Diagenode kit were used as recommended by the manufacturer. The longer immunoselection step (15 hours over 5 hours) at 4°C was selected for the auto MeDIP reactions ("MeDIP_19h_DIB" program) and the recommended 10 hour immunoseletion step at 4°C was selected for the auto hMeDIP reactions ("hMeDIP_14h_DIB") on the IP-Star. After the incubations were complete, all samples were removed from the IP-Star and the tubes containing either the immunoprecipitated sample DNA or the input DNA were placed in a magnetic rack to capture the DNA-bound beads for a minimum of 15 minutes. The supernatant fractions were transferred to new tubes and labelled as either 'input' or 'IP' as appropriate before proceeding to the first quality control step (QC1). 4.5µl of each DNA sample was set aside for QC1 and the remaining DNA was purified using Agencourt Ampure XP beads and eluted in 22µl buffer EB.

2.8.5 Quality control (part 1)

DNA recovery of hydroxymethylated, methylated and unmethylated spiked-in DNA and specificity of the antibody immunoprecipitation was tested using quantitative PCR (qPCR). The primer sequences used for QC1 of MeDIP samples (corresponding to the '20CpG' and '15CpG' lambda DNA fragments) are listed in Supplementary Table 1. The primer used for QC1 of hMeDIP samples were provided in the auto hMeDIP kit. 4.5µl of each DNA library and its respective input DNA was diluted to 10µl in ultra-pure H_2O. Triplicate qPCR reactions were set up on ice in 96-well plates for each DNA library as follows: 6.25µl MESA BLUE qPCR MasterMix Plus for SYBR (Eurogentec; #RT-SY2X-03+WOUB), 0.625µl forward primer (10µM), 0.625µl reverse primer (10µM), 3.75µl ultra-pure H_2O, and 1.25µl template DNA. The qPCR plates were run on a 7300 Real Time PCR System (Applied Biosystems) on the following program: 95ºC for 15min, 40 cycles of 95ºC for 15s and 60ºC for 1min, followed by melt curve analysis. The percentage recovery for each spiked-in DNA fragment was calculated as follows:

$$Recovery\ (\%) = 2AE^{(Adjusted\ input\ Ct - MeDIP\ Ct)} \times 100$$

AE refers to the amplification efficiency of the primer pair (e.g. 100% = 1) and C_t represents the cycle threshold calculated from the qPCR results. The MeDIP C_t refers to the cycle threshold calculated for the sample DNA (either MeDIP or hMeDIP libraries). The C_t value of the input DNA must be adjusted to reflect the differing amounts of DNA that are included for input compared to the IP reactions, as only 10% of the total library-prepped DNA sample is used as input. This is achieved as follows:

$$Adjusted\ input\ Ct = Input\ Ct - \log\left[2AE(10)\right]$$

Finally, specificity of the antibody towards its respective modification (either 5mC or 5hmC) was calculated as follows:

$$Specificity = 1 - \frac{recovery_{unmethylated}}{recovery_{methylated}}$$

Successfully prepared MeDIP and hMeDIP libraries all had a specificity ≥ 95%.

2.8.6 Library preparation (part 2)

Immunoprecipitated DNA samples underwent adapter-mediated PCR amplification in order to generate suitable libraries for next-generation sequencing using the 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems; #KK2601) and the primers provided by the Multiplexing Sample Preparation Oligonucleotide Kit (Illumina; #PE-400-1001). 22µl of purified MeDIP or hMeDIP DNA was used as template DNA in a final reaction volume of 50µl. A PCR reaction was set up for each library as follows: 1µl of PCR Primer InPE 1.0 (25µM), 1µl of PCR Primer InPE 2.0 (0.5µM), 1µl of PCR Primer Index (25µM), 25µl of KAPA HiFi HotStart ReadyMix and 22µl template DNA. The PCR reaction was incubated on a thermal cycler under the following conditions: 98ºC for 45s, 15 cycles of 98ºC for 15s, 65ºC for 30s, 72ºC for 30s, followed by a final extension at 72ºC for 5min and hold at 4ºC. DNA was purified using Ampure XP beads and eluted in 18µl buffer EB.

2.8.7 Library size selection

PCR-amplified, purified MeDIP and hMeDIP DNA samples were loaded onto a 2% agarose gel and analysed by gel electrophoresis (1hr 40min at 100V). 3 wells were left empty on either side of each sample to prevent cross-contamination during gel excision. In addition, each sample was separated by a 50bp DNA ladder (New England Biolabs) in order to facilitate accurate library size selection. The gel was transferred to a UV transilluminator and a gel band corresponding to 300-350bp was excised for each DNA library and collected in a 2ml DNA LoBind tubes (Eppendorf; #0030108.078). In addition, bands corresponding to 250-300bp and 350-400bp DNA size ranges were also excised to serve as back-up libraries. The DNA was extracted from each gel band using the MinElute Gel Extraction Kit (Qiagen; #28604) according to the manufacturer's protocol, using an overnight incubation in Buffer QG at 4ºC in order to ensure full dissolution of the gel slice. DNA was eluted in 10µl buffer EB.

2.8.8 Quality control (part 2)

The fold-enrichment ratio of hydroxymethylated and methylated DNA fragments present within the hMeDIP and MeDIP DNA libraries, respectively, was calculated using primers to assay known hydroxymethylated, methylated and unmethylated regions in the endogenous genomic DNA using qPCR. Primer pairs used for hMeDIP QC2 were provided in the Auto hMeDIP kit (Diagenode; #AF-Auto02-0016). MeDIP QC2 primer pairs are listed in Supplementary Table 1; the 5mC status of these two amplified regions had been previously validated in mESC DNA by past lab members. 1µl of size-selected and purified DNA was diluted to 10µl using ultra-pure H_2O for use as template DNA in the qPCR reactions. All qPCR reactions were assayed in triplicate as described in section 2.8.5. The fold-enrichment ratio was calculated for two regions per DNA library (methylated and unmethylated, or hydroxymethylated and unmethylated) as follows:

$$Fold-enrichment\ ratio = \frac{2AE^{(Ct_{Input\_meth} - Ct_{MeDIP\_meth})}}{2AE^{(Ct_{Input\_unmeth} - Ct_{MeDIP\_unmeth})}}$$

where AE represents the amplification efficiency of the primer pair in question, expressed as a number (e.g. 100% efficiency = 1). Fold-enrichment ratios ≥ 25 were considered suitable for progressing with DNA sequencing (Taiwo et al., 2012).

2.8.9 DNA sequencing

MeDIP and hMeDIP libraries were subjected to next-generation DNA sequencing using 100 base pair paired-end reads on an Illumina HiSeq 2000 according to manufacturer's instructions by the sequencing facility at Illumina (Cambridge, UK).

2.8.10 Sequencing data analysis

Analysis of the raw sequencing fastq data files was performed by Dr Gareth Wilson (Medical Genomics, UCL Cancer Institute, University College London) using the MeDUSA analysis pipeline (Wilson et al., 2012). Briefly, the MeDUSA pipeline was used to align the sequencing reads to the reference genome mm10 using the Burrows-Wheeler alignment tool (BWA), followed by filtering of clonal and unpaired reads with SAMtools. 5mC and 5hmC peaks were defined using MACS

(q>10) (Zhang et al., 2008) prior to DMR calling between biological cohorts using 500bp windows.

2.9 Infinium HumanMethylation450 BeadChip

2.9.1 Bisulphite conversion

Bisulphite (BS) conversion of Saos-2 cell line DNA samples for 450K methylation profiling (see Chapter 5, sections 5.3.4 and 5.3.6) was performed using the EZ DNA Methylation Kit (Zymo Research; #D5001) according to manufacturer's instructions. 500ng of input genomic DNA was used per conversion and all samples were incubated with BS reagent using the thermal cycling conditions specifically recommended for the Illumina Infinium 450K analysis, consisting of 16 cycles of 95°C for 30s plus 50°C for 60min, followed by holding samples at 4°C until DNA purification. DNA was eluted in a final volume of 12µl and stored at -80°C until BeadChip processing.

2.9.2 Bisulphite conversion efficiency

BS conversion efficiency was assessed for all Soas-2 DNA samples prior to 450K analysis (see Chapter 5, section 5.3.6) in order to confirm successful, complete conversion of the DNA sample. Conversion efficiency was assessed by qPCR using two primer pairs designed around the same 133bp sequence located at chr7:5571729-5571861 (hg19 assembly), approximately 1kb upstream of the *ACTB* gene. One primer pair ("Actin positive") specifically amplifies BS-converted DNA only; the second primer pair ("Actin negative") specifically amplifies unconverted DNA only (see Supplementary Table 1 for primer sequences). Each DNA sample was assayed by qPCR in triplicate for both primer pairs. All qPCR reactions were set up in a final volume of 12.5µl as follows: 6.25µl MESA Blue qPCR MasterMix Plus for SYBR (Eurogentec; #RT-SY2X-03+WOUB), 0.625µl forward primer (10µM), 0.625µl reverse primer (10µM), 3.75µl ultra-pure H_2O, and 1.25µl template DNA. The qPCR plates were run on a 7300 Real Time PCR System (Applied Biosystems) on the following program: 95ºC for 10min, 40 cycles of 95ºC for 15s and 60ºC for 1min, followed by melt curve analysis. BS conversion efficiency was calculated as follows:

$$\Delta Ct = Mean\ Ct^{Actin\ Negative} - Mean\ Ct^{Actin\ Positive}$$

$$Conversion\ efficiency\ (\%) = \left[1 - \left(\frac{1}{1 + 2^{\Delta Ct}}\right)\right] \times 100$$

A conversion efficiency of ≥ 99% was obtained for all samples.

2.9.3 Oxidative bisulphite conversion

Oxidative bisulphite (oxBS) conversion of human brain DNA samples was performed using either a "low input" or "high input" protocol (see Chapter 5, section 5.2.4).

2.9.3.1 Low-input protocol

1µg DNA per sample was processed using a trial version of the TrueMethyl 24 kit (CEGX). Samples were split evenly into two aliquots of 500ng and processed through either the BS-only or oxBS conversion workflow. Each aliquot was subjected to an initial buffer exchange step using a spin column format and eluted in ultra-pure water. The full eluate (~22µl per sample) was carried forward. Samples were denatured using the provided denaturing solution for 30 min at 37 °C, in a total reaction volume of 24µl and immediately taken forward to oxidation. 1µl of the provided oxidant solution was added to each sample undergoing the oxBS workflow only (1µl of ultra-pure water was added to BS-only samples for mock oxidation). All samples were incubated for 30 min at 40 °C. Oxidised samples were equilibrated to room temperature before proceeding immediately to BS conversion. BS reagent solution was prepared as described in the TrueMethyl protocol and 170µl was added to each 25µl oxidation reaction mixture. Finally, 5µl of BS additive was added to each reaction, bringing the total volume to 200µl. All reactions were incubated using BS-specific thermal cycling conditions. Converted DNAs were purified using the provided spin columns and 4µl of each sample eluate was used as input into the Infinium 450K array.

2.9.3.2 High-input protocol

4µg DNA per sample was processed using the TrueMethyl protocol for 450K analysis (Version 1.1, CEGX). The protocol was performed as described, with two minor alterations. First, the initial shearing step of DNA was performed using a Bioruptor (Diagenode). The DNA was diluted into a total volume of 50µl in 1.5 ml microtubes and sonicated for 30 seconds on setting 'H'. 10µl was set aside for visualisation on a 2% agarose gel to confirm correct DNA fragmentation (all <10 kb). The remaining 40µl (~3.2µg) sheared DNA was then split into 2 × 20µl aliquots and purified using Agencourt AMPure XP beads using a modified protocol described in TrueMethyl 24 Kit User Guide (Appendix 3, version 3.1, CEGX). Each sample was eluted in 20µl ultra-pure water and quantified using a NanoDrop 1000 spectrophotometer. The the spike-in CEGX digestion control was added at a level of 0.5% w/w and then DNA samples were processed either through the oxBS or BS-only workflow as outlined in the TrueMethyl 24 Kit User Guide. 7µl of the final eluate was used as input into the Infinium 450K array, made possible by adding 1µl of 0.4 N NaOH to each sample rather than the standard 4µl of 0.1 N NaOH when setting up the MSA4 plate during BeadChip processing.

2.9.4 Taq1 digestion control

For assessment of the extent of oxBS conversion, the spike-in digestion control was PCR amplified from both oxBS and BS-only DNA conversions and subjected to Taq1 restriction endonuclease digestion. The digestion control contains a single Taq1 recognition site (5'-TCGA-3') with 5hmC in the internal CpG site. After successful oxBS conversion and PCR amplification, the internal 5hmC should have undergone complete conversion to T, rendering the amplicon resistant to Taq1 digestion. The BS-converted and PCR-amplified digestion control should retain the internal CpG site and therefore undergo cleavage by Taq1.

PCR amplification of the BS- and oxBS-converted DNA (see section 2.7.2) was carried out with DreamTaq DNA polymerase (Thermo Scientific; #EP0701). Each reaction was set up as follows: 1µl DNA as template, 60µl 10X DreamTaq Buffer, 24µl dNTP mix (10mM), 6µl digestion control forward primer (100µM), 6µl

digestion control reverse primer (100µM), 480µl ultra-pure H_2O, and 12µl DreamTaq DNA polymerase (5U/µl). All reactions were incubated on a thermal cycler under the following conditions: 95°C for 120s followed by 40 cycles of 95°C for 30s, 60°C for 30s, and 72°C for 15s. PCR reactions were purified using the QIAquick PCR Purification Kit (Qiagen; #28104) and eluted in 50µl EB (provided with kit). Purified DNA was quantified using a NanoDrop 1000 spectrophotometer and concentration adjusted to 10ng/µl using ultra-pure H_2O.

Each DNA sample was subsequently digested by the Taq1 restriction endonuclease (Promega; #R6151) in a final reaction volume of 20µl containing: 10µl of template DNA (10ng/µl), 2µl of 10X RE Buffer, 4µl of Taq1 (10U/µl), 0.2µl BSA (10µg/µl) and 3.8µl ultra-pure H_2O. Positive and negative control reactions were carried out in parallel, with either 5µl of cutting control DNA (20ng/µl; provided by TrueMethyl kit) replacing the amplified DNA sample or without the addition of Taq1 DNA polymerase. Each control reaction was made up to a final volume of 20 µl with additional ultra-pure H_2O. All digestion reactions were incubated on a thermal cycler at 65°C for 18 hours followed by an inactivation step at 80°C for 20 minutes. Reaction products were analysed by gel electrophoresis using a 2% agarose gel.

2.9.5 BeadChip processing

BS-converted and oxBS-converted DNA samples were interrogated on Illumina Infinium HumanMethylation450 BeadChips following the manufacturer's recommendations by UCL Genomics (Institute of Child Health Microarray and Next Generation Sequencing Facility, University College London). This required the use of 4µl of eluted DNA and 4 µl of 0.1N NaOH per sample. However, an optimisation of the TrueMethyl protocol, referred to as the "high-input" protocol in Chapter 5 (section 5.2.4), required that oxBS-converted samples were processed on the 450K BeadChip using 7µl of eluted DNA and 1 µl of 0.4N NaOH per sample in order to maximise the amount of DNA interrogated on the arrays. This applied to a total of four samples (2 BS-converted replicates and 2 oxBS-converted replicates of the "brain1" DNA sample).

2.9.6 Data analysis

The human brain samples subjected to the oxBS protocol (see Chapter 5, section 5.2) were analysed independently of the Saos-2 cell line dataset (see Chapter 5, section 5.3) processed under a standard BS conversion protocol. Raw IDAT files were obtained from UCL Genomics and processed using ChAMP, a 450K analysis pipeline implemented with the statistical software package R (v.3.0.1) (Morris et al., 2014). Probes were filtered by their raw intensity values using a detection p-value threshold of 0.01 and normalised using SWAN (Makismovic et al., 2012). For the human brain samples, an additional probe-filtering step was implemented: probes corresponding to the X and Y chromosomes were removed from the dataset as both male and female samples were being analysed.

2.10 Pyrosequencing

DNA underwent oxBS conversion using the TrueMethyl 24 Kit followed by pyrosequencing analysis ("oxBS-pyroseq") on a PyroMark Q96 MD system (Qiagen) to quantify both 5mC and 5hmC at individual CpG sites.

2.10.1 Primer design

A subset of differentially methylated regions called from MeDIP-seq and hMeDIP-seq datasets were selected for validation based their statistical significance. Regions falling within repetitive DNA sequence were excluded because they could not be uniquely targeted and amplified through PCR. For oxBS-450K validation, 50bp windows containing at least two CpG sites identified as containing the highest levels of 5hmC in the 'brain1' sample (5hmC $\geq$ 30%) were selected for oxBS-pyroseq.

PyroMark Assay Design software (v.2.0.1.15) was used to design and select the most suitable primers for each pyrosequencing assay. Final primer lists used for pyrosequencing analysis in validation experiments are detailed in Supplementary Table 1. All custom-designed primers were ordered from Sigma-Aldrich.

2.10.2 Primer standard curves

Methylated DNA standard curves for each primer pair were generated to check for primer bias due to the 5mC content of DNA samples. A control DNA sample (pooled

human blood sample; see section 2.2) was used to generate fully unmethylated and fully methylated DNA aliquots that were then mixed at differing ratios to create pools of 0, 20, 40, 60, 80 and 100% methylated DNA.

Fully unmethylated DNA was generated by undergoing two successive rounds of whole genome amplification (WGA) using the REPLI-g Midi Kit (Qiagen; #150043), according to the manufacturer's protocol. The WGA reaction mix was incubated for 16 hours at 30°C in order to maximise the DNA yield.

Fully methylated DNA was generated by treatment with M.SssI, a CpG methyltransferase (New England Biolabs; # M0226L). *In vitro* methylation reactions were set up in a final volume of 50µl as follows: 5µl 10X NEBuffer 2, 5µl SAM (1600µM), 10µl template DNA (300ng/µl), 27.5µl ultra-pure H_2O and 2.5µl M.SssI (4U/µl). Reaction mixes were incubated on a thermal cycler for 1 hour at 37°C followed by 20 minutes at 65°C to inactivate the enzyme.

The methylation status of the fully unmethylated and methylated control DNA samples were assessed through digestion with three different methylation-sensitive restriction endonucleases: HpaII (New England Biolabs; #R0171L), HpyCH4IV (New England Biolabs; # R0619L) and BstUI (New England Biolabs; #R0518L). Digestion reactions were set up for each control DNA sample in a final volume of 40µl as follows: 10µl control DNA (~50ng/µl), 24µl ultra-pure H_2O, 4µl NEBuffer1, 1µl HpaII (10U/µl) and 1µl HpyCH4IV (10U/µl). Negative control reactions without either enzyme were also set up for each control DNA sample. All reactions were incubated on a thermal cycler for 1 hour at 37°C, before adding 1µl BstUI (10U/µl) to each (with the exception of negative controls) and incubating for a further 1 hour at 60°C. After the incubation, reaction mixes were analysed on a 2% agarose gel to check for digestion. Only WGA control DNA was visibly digested, whilst M.SssI-treated DNA and both negative controls remained as a high molecular weight, tight DNA band (>10kb) after gel electrophoresis.

The remaining WGA and M.SssI-treated DNA were purified using Agencourt Ampure XP beads, eluted in 100µl buffer EB (Qiagen) and the concentration adjusted to 50ng/µl each and checked using a Qubit dsDNA broad range assay kit

(Invitrogen; #Q32853). Next, six different pools of 0, 20, 40, 60, 80 and 100% methylated DNA were made up by mixing differing ratios of WGA and M.SssI-treated DNA in a final volume of 20µl (i.e. 1µg DNA per mixture) and then subjecting each sample to BS conversion using the EZ DNA Methylation Kit (Zymo Research; ##D5001), according to the manufacturer's protocol. All samples were eluted in a final volume of 40µl M-elution buffer and adjusted to a 10ng/µl final concentration.

PyroMark PCR reactions for each primer pair (see section 2.10.1) were set up in triplicate on the six different BS-converted, control DNA samples (0, 20, 40, 60, 80 and 100% 5mC), each in a final volume of 25µl and set up as follows: 23µl of a PCR mastermix (12.5µl PyroMark PCR mastermix, 2µl $MgCl_2$ (25mM), 2.5µl 10X CoralLoad, 6µl H_2O), 1µl DNA template, 0.5µl forward primer (5µM) and 0.5µl reverse primer (5µM). PCR plates were incubated on a thermal cycler under the following conditions: 95°C for 15min, 13 cycles of 94°C for 30s, 66°C for 30s (-0.5°C per cycle) and 72°C for 30s, followed by 50 cycles of 94°C for 30s, 56°C for 30s, 72°C for 30s, and then a final extension of 72°C for 10min.

Finally, each PCR product was subject to pyrosequencing analysis (in triplicate) on a PyroMark Q96 MD system, according to the manufacturer's protocol (see section 2.10.6 for further details). Standard curves were generated for each primer pair by plotting expected methylation scores against pyrosequencing-derived measurements of methylation score (Supplementary Figure 1) and the final pyrosequencing results adjusted based on correlation between observed and expected 5mC scores.

2.10.3 DNA extraction

DNA was purified from *in vitro* cultured cell samples for oxBS-pyroseq analysis using the DNeasy Blood & Tissue Kit (Qiagen; #69504), according to manufacturer's instructions.

2.10.4 Oxidative bisulphite conversion of DNA

4µg of each DNA sample was diluted in ultra-pure H_2O to a final volume of 150µl and sonicated for 30 seconds using a Biorupter (Diagenode) on the "high" setting.

Each sample was then purified using the Zymo Clean and Concentrator-25 kit, following the manufacturer's instructions, and eluted in 45µl ultra-pure H_2O. Purified DNA was quantified using a NanoDrop 1000 spectrophotometer and the spike-in CEGX digestion control was added to each sample at a level of 0.5% w/w. Each sample was then split into two aliquots of 22.5µl and processed through the high-input TrueMethyl protocol as described previously (see section 2.9.3).

2.10.5 PCR amplification

PCR reactions were set up in 96-well plates using the PyroMark PCR kit (Qiagen; #978703). Each PCR was run in triplicate on both oxBS- and BS-converted DNA for each primer pair. PCR reactions were carried out in a final volume of 25µl and contained 23µl of a PCR mastermix (12.5µl PyroMark PCR mastermix, 2µl $MgCl_2$ (25mM), 2.5µl 10X CoralLoad, 6µl H_2O), 1µl DNA template, 0.5µl forward primer (5µM) and 0.5µl reverse primer (5µM), one of which was biotinylated on the 5' end (see section 2.10.1). PCR plates were incubated on a thermal cycler under the following conditions: 95°C for 15min, 13 cycles of 94°C for 30s, 66°C for 30s (-0.5°C per cycle) and 72°C for 30s, followed by 50 cycles of 94°C for 30s, 56°C for 30s, 72°C for 30s, and then a final extension of 72°C for 10min.

2.10.6 Sequencing run

PCR amplicons were prepared for sequencing through immobilisation to streptavidin- and sepharose-coated agarose beads. Each reaction was set up in a final volume of 70µl within a 96-well plate format ("plate 1"). Each well contained: 10µl of biotinylated PCR product, 40µl PyroMark binding buffer (Qiagen; #979006), 28µl ultra-pure H_2O and 2µl streptavidin sepharose high performance agarose beads (GE Healthcare; #17-5113-01). The plate was sealed with a polyolefin film (Starlab; #E2796-9795) and agitated for 10 minutes on a microplate shaker (SciQuip; #SP2260-MP).

Next, a 96-well PyroMark Q96 plate (Qiagen; #979101), referred to as plate 2, was set up containing the sequencing primers corresponding to each of the biotinylated PCR amplicons present in plate 1. First, 10µM sequencing primer stocks were diluted to 0.3µM in PyroMark annealing buffer (Qiagen; #979009) and then 12µl of

each sequencing primer was dispensed into the appropriate well of plate 2, taking care not to scratch the lower surface of the well.

The PyroMark Q96 Vacuum Workstation was set up according to the manufacturer's recommendations for capture of the amplicon-bead complexes in plate 1, followed by several wash steps and denaturation (using 0.2M NaOH) in order to generate single-stranded DNA fragments suitable for pyrosequencing. Finally, the single-stranded DNA fragments were dispensed into plate 2 in order to mix with their corresponding sequencing primer. Plate 2 was sealed and placed in a heating block at 85ºC for 5 minutes before cooling to room temperature, in order to ensure primer annealing to the template DNA strands.

Pyrosequencing assays were analysed on a PyroMark Q96 MD system and the data output analysed using Pyro Q-CpG software (v.1.0.9). Methylation scores at each CpG site were normalised against a linear calibration curve (see section 2.10.2). 5mC levels were obtained from the oxBS-converted methylation scores, whilst 5hmC levels calculated by subtraction of oxBS from BS-only methylation scores.

2.11 LC-MS analysis

Liquid chromatography-mass spectrometry (LC-MS) analysis was performed by Martin Bachman (Department of Chemistry, University of Cambridge) as follows: 500ng of genomic DNA was incubated with 5 U of DNA Degradase Plus (Zymo Research; #E2020) at 37°C for 3 hours. The resulting mixture of 2'-deoxynucleosides was analysed on a Triple Quad 6500 mass spectrometer (AB Sciex) fitted with an Infinity 1290 LC system (Agilent) and an Acquity UPLC HSS T3 column (Waters), using a gradient of water and acetonitrile with 0.1% formic acid. External calibration was performed using synthetic standards, and for accurate quantification, all samples and standards were spiked with isotopically labelled nucleosides.

Chapter 3 Development of an *in vitro* culture and phenotyping model of neural differentiation

3.1 Introduction

The central aim of this project was to investigate the potential function of 5hmC through profiling of both 5mC and 5hmC in parallel during the differentiation of mouse embryonic stem cells (mESCs) down a neural lineage. Unlike 5mC, global 5hmC levels vary widely between somatic tissue types and are highest in the CNS, where it has been found in the range of 0.3 to 0.8% of all cytosines (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Munzel et al., 2010; Nestor et al., 2011). Lower but nevertheless significant levels of 5hmC have also been observed in mouse ESCs, ranging from 0.032% (Tahiliani et al., 2009) to 0.047% of total nucleotides (Song et al., 2011). In addition to non-neural somatic tissues, the lowest levels of 5hmC are observed in cells with high proliferative capacity, including blood, immortalized cell lines such as HeLa, HEK293FT and HCT116 and numerous cancers (Haffner et al., 2011; Li and Liu, 2011; Song et al., 2011). Even if present at very low global levels, 5hmC may still have biological relevance if it shows a clustered, rather than random, distribution throughout the genome, such as marking hotspots of DNA demethylation activity.

An association between differentiation stage and 5hmC level has also been noted through immunohistochemical staining of stratified epithelial cells in various embryonic and adult tissues, where undifferentiated basal cells showed reduced 5hmC staining as compared to differentiated apical cells (Haffner et al., 2011). The apparent inverse correlation between 5hmC levels and proliferative capacity led to the hypothesis that the most significant changes to 5hmC distribution (in the form of 5hmC gains) may occur during normal CNS development, as pluripotent stem cells undergo lineage commitment to terminally differentiated neural cells.

Although high levels of 5hmC had been measured in various mouse tissue types of the CNS, including cortex, hippocampus, cerebellum and hypothalamus, there was little information available on variations in 5hmC distribution amongst neural cell subtypes. Purkinje and granule cell neurons have been shown to contain some of the highest recorded levels of 5hmC at 0.6 and 0.2% of all nucleotides, respectively, but

comparable data for either astrocytes or oligodendrocytes was not yet available (Kriaucionis and Heintz, 2009). However, it seemed unlikely that neurons alone could be responsible for the high 5hmC signal observed in CNS tissues. The mouse neocortex, for example, has a glial-to-neuronal ratio of approximately 3:1 (Fu et al., 2012) and contains the highest global 5hmC levels of any tissue. On the other hand, neurons outnumber glial cells by approximately 4:1 in the mouse cerebellum (the most neuron-dense tissue of the CNS) and yet 5hmC levels are up to 50% lower here than in cortical tissue (Globisch et al., 2010; Munzel et al., 2010; Szwagierczak et al., 2010), suggesting that glial cells are likely to contribute to global 5hmC levels to some degree. In the absence of any concrete data on the differences in 5hmC content between neurons and glia, the relative ease of *in vitro* differentiation of the various neural subtypes was taken into account in order to determine a suitable cell culture model for further analysis.

Fully differentiated neural cells may be derived in culture from either ESCs (via an intermediary neural precursor stage) or from primary neural stem cells (NSCs) isolated from the embryonic or adult brain (Figure 3-1). Neural precursors may be maintained in culture as either floating neurospheres or as an adherent monolayer upon addition of growth factors. The first example of isolated primary NSCs were propagated in culture as floating neurospheres (Reynolds and Weiss, 1992), heterogeneous spherical clusters of cells comprising of neurons, oligodendrocytes and astrocytes in addition to neural progenitors (Bez et al., 2003; Suslov et al., 2002). It is thought that the differentiated neural cells provide a cellular niche that allows for the maintenance of an NSC subpopulation within the neurosphere.

More recently, successful expansion of ESC-derived NSCs in adherent conditions has been demonstrated, resulting in a homogenous population of cells with stable, long-term properties of self-renewal and tripotent differentiation capacity (Conti et al., 2005; Glaser et al., 2007; Ying et al., 2003). Alternatively, primary neural stem cells may be isolated from the subventricular zone (SVZ) of the foetal adult brain and established in adherent monoculture under the same conditions as embryonic-derived NSCs, retaining the capacity to produce differentiated neuronal and glial

cells throughout many passages (Conti et al., 2005; Kriegstein and Alvarez-Buylla, 2009; Pollard et al., 2006).

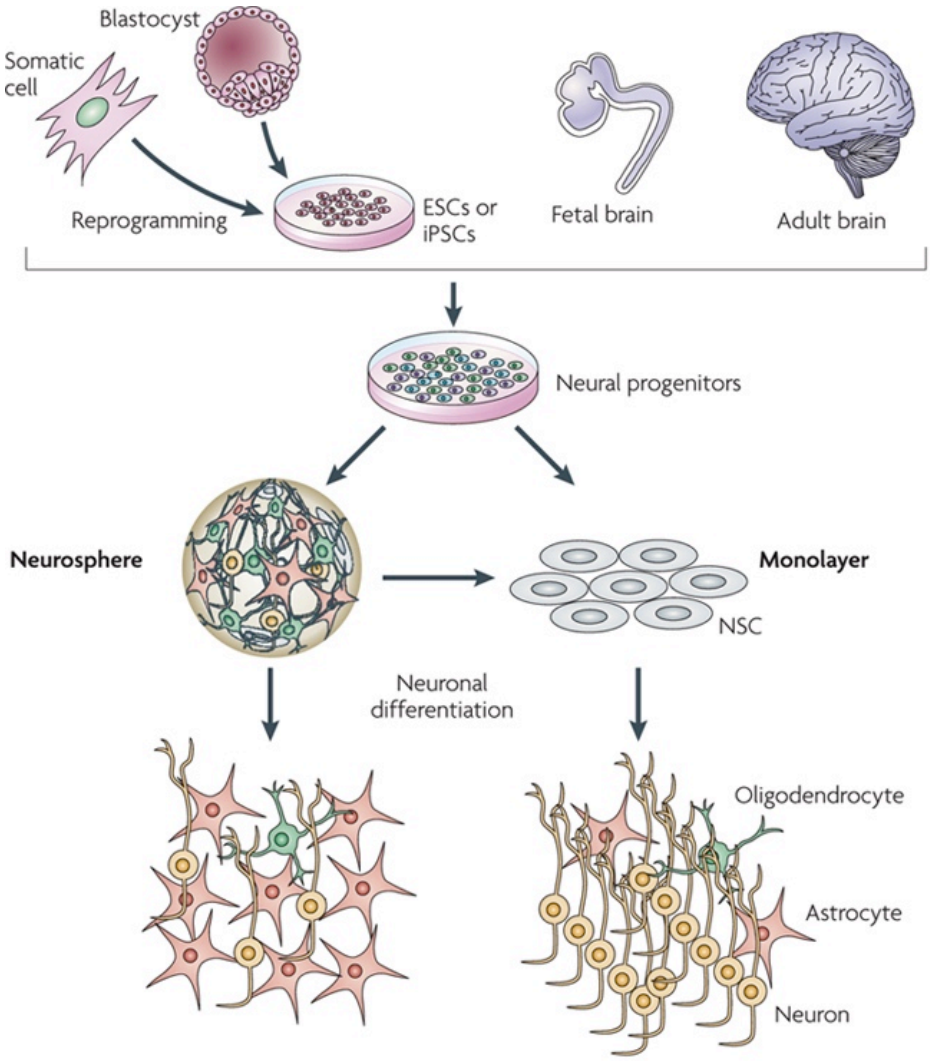


Figure 3-1: Overview of *in vitro* neural stem cell systems. Neural progenitors may be derived through the *in vitro* differentiation of ESCs and iPSCs or isolated from foetal and adult brain. Neural progenitors may be expanded within neurospheres (a heterogeneous cell mass) or as a homogeneous, adherent monolayer before undergoing further neural lineage differentiation. Image courtesy of: Conti L, Cattaneo E. 2010. *Nature Reviews Neuroscience* **11**, 176-187**.**

The astrocytic differentiation of neural stem cells is well established and produces homogeneous cultures of functional, mature astrocytes after just 48 hours of exposure to serum (Brunet et al., 2004; Sauvageot and Stiles, 2002). These cells uniformly express glial fibrillary acidic protein (GFAP), a type of intermediate filament protein present in mature astrocytes (Jacque et al., 1978), generally accepted as the prevailing immunocytochemical (ICC) marker of this cell type and whose expression is induced by serum exposure (Eng et al., 2000; Sakai et al., 1990).

Although many protocols for the *in vitro* production of neurons are available, they are technically challenging, difficult to reproduce and result in highly variable yields of neuronal cells. Neuronal differentiation is often accompanied by high rates of cell death and results in heterogeneous cell cultures comprising both neurons and glial cells, requiring further cell sorting or lineage selection to isolate a particular neural subtype (Gaspard et al., 2008; Kim et al., 2002; Ying and Smith, 2003). More recently, efficient generation of neurons in high yield (90% or more) have been reported from adherent, monolayer mESC cultures, although these neurons display heterogeneity in both marker expression and physiological function (Li et al., 2009).

The *in vitro* generation of oligodendrocytes from either ESCs or NSCs has traditionally proved more difficult to achieve. An initial protocol for the generation of differentiated glial cells expressing the oligodendrocyte-specific markers O4 and 2′, 3′-cyclic nucleotide 3′-phosphodiesterase (CNP) from rat ESCs reported heterogeneous populations comprising of approximately equal percentages (30-40%) of oligodendrocytes and astrocytes (Brüstle et al., 1999). Others have employed floating 'oligosphere' cultures to produce > 90% mature oligodendrocytes from mESCs. This multi-step procedure involves a combination of embryoid body (EB) formation, induction with retinoic acid and transfer into a defined oligosphere medium to form free-floating cell clusters that may dissociated and replated to encourage differentiation into mature oligodendrocytes (Liu et al., 2000). More recently, differentiated oligodendrocytes have been derived in adherent monolayers using a stably transfected CNP-ßgeo mESC line followed by a neomycine selection step to produce glial precursors that could be further differentiated into galactocerebroside-positive oligodendrocytes at yields of 80 ± 7% (Glaser et al., 2005). A simpler protocol starting from a stably renewing mNSC population has also been demonstrated, but results in heterogeneous populations of neural cells comprising of approximately 20% oligodendrocytes, 40% astrocytes and 10% neurons (Glaser et al., 2007).

Astrocytes were therefore deemed to be the most suitable differentiated cell type for analysis, as large numbers of homogeneous cells may be easily and efficiently produced through *in vitro* differentiation of mNSCs. In order to elucidate the precise

changes that occur in genomic 5hmC and 5mC distribution along the neural differentiation pathway, a combined MeDIP- and hMeDIP-seq approach was used to profile the DNA methylome at three defined cellular stages: mESCs, mNSCs and fully differentiated astrocytes. This chapter describes the *in vitro* cell culture model used to generate these cells for 5mC/5hmC DNA profiling.

3.2 Establishment of mouse embryonic stem cells in culture

JM8, a wild-type (WT) C57BL/6N mESC line, was initially established in culture using a LIF- and serum-containing growth medium in the absence of feeder cells (Pettitt et al., 2009; Smith, 1991). However, considerable spontaneous differentiation was observed under these growth conditions (Figure 3-2A), which did not improve during subsequent passaging.

It has previously been observed that mESCs cultured in traditional LIF- and serum-containing medium exhibit cellular heterogeneity in transcription factor expression (Chambers et al., 2007; Hayashi et al., 2008; Payer et al., 2006; Toyooka et al., 2008). Several positive regulators of pluripotency (including NANOG, STELLA, OCT4 and KLF4) are present in varying concentration amongst subpopulations of mESCs, resulting in differences in self-renewal and differentiation capacity. These subpopulations appear to represent interconvertible cell states resulting from continuous, stochastic fluctuations in transcription factor expression and associated epigenetic modifications. This presents a problem for studies employing epigenetic and gene expression profiling methods, potentially confounding any downstream comparative analysis performed on datasets derived from such samples.

In order to improve the homogeneity of the mESC culture and ensure they represented a truly undifferentiated population of cells, an alternative '3i' culture method was tested. '3i' refers to LIF- and serum-free growth media that relies on the addition of three small-molecule inhibitors SU5402, PD184352 and CHIR99021 to block mESC differentiation (Ying et al., 2008). These inhibitors target FGF receptor tyrosine kinases, the mitogen-activated protein kinase (MAPK) pathway and glycogen synthase kinase-3 (GSK3), respectively. Under these conditions, mESCs express uniformly high levels of *Nanog*, *Klf4* and *Rex1* and are thought to represent a

developmentally neutral, 'ground state' of pluripotency, maintaining the capacity for continuous self-renewal and tripotent differentiation (Wray et al., 2010).

When re-plated in 3i medium (iSTEM®, StemCells Inc.; #SCS-SF-ES-01), the JM8 mESCs exhibited an immediate change in cellular morphology. Cells grew as uniform colonies with little or no visible differentiation (Figure 3-2B).

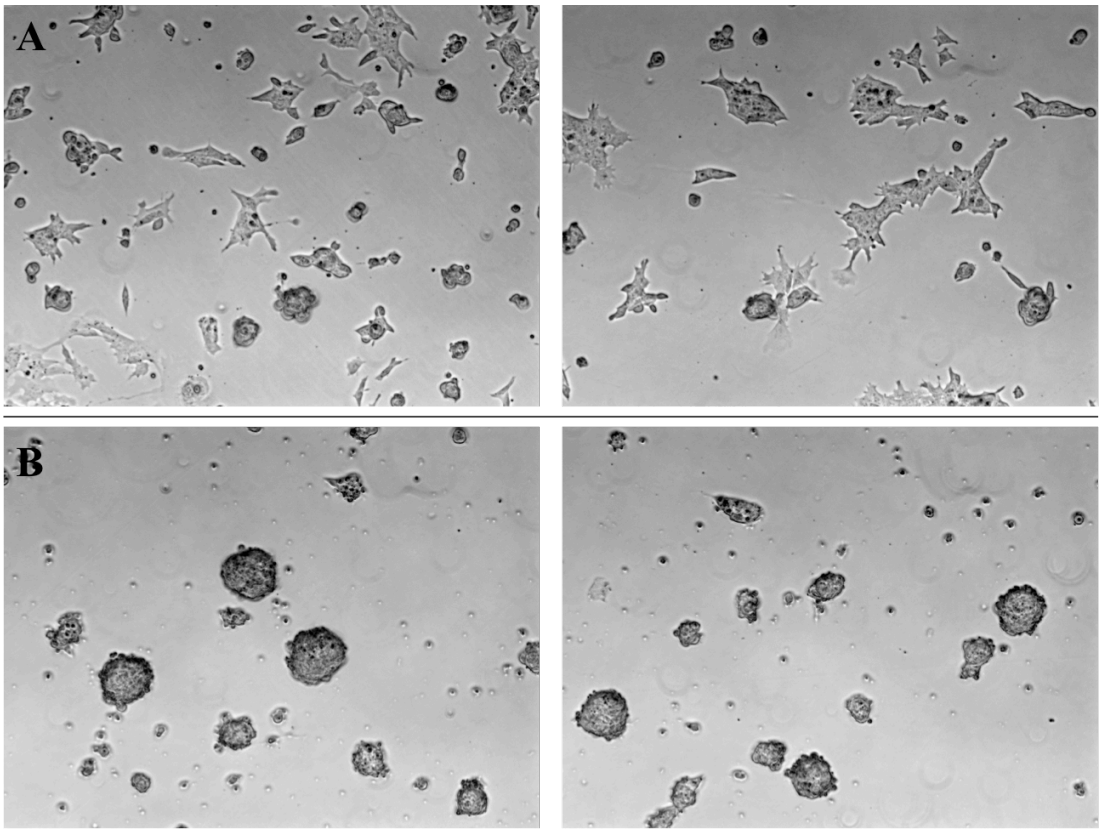


Figure 3-2: Phase contrast images of JM8 mESCs grown in adherent conditions in (A) LIF- and serum-containing media and (B) 3i media. In both conditions shown above, mESCs are plated on gelatin-coated plasticware. Significant numbers of differentiated cells are observed in condition (A), characterised by flat, irregularly shaped morphology and large cytoplasm. mESC colonies, in contrast, are characterised by small, compact cells with large nucleus-to-cytoplasm ratios. Colonies are usually circular and have a slightly raised, dome-shaped appearance.

3.2.1 Embryonic stem cell phenotyping

JM8 cells grown in either 3i or LIF- and serum-containing media underwent paraformaldehyde (PFA) fixation and subsequent immunostaining using an anti-OCT-3/4 monoclonal antibody (Santa Cruz Biotechnology, Inc.; #sc-5279), a well-established marker of ESCs (Nichols et al., 1998; Niwa et al., 2000). Uniform, positive *Oct4* expression was observed for all mESC colonies kept in both growth

media conditions (Figure 3-3A and B), although several isolated OCT4-negative cells were observed surrounding the mESC colonies (indicated by DAPI nuclear counterstaining) when grown in LIF and serum, indicating the presence of some differentiating cells. No immunoreactivity against the neural progenitor cell markers nestin and SOX1 was detected, providing confidence that these cells were truly ESCs (Figure 3-3C-D).

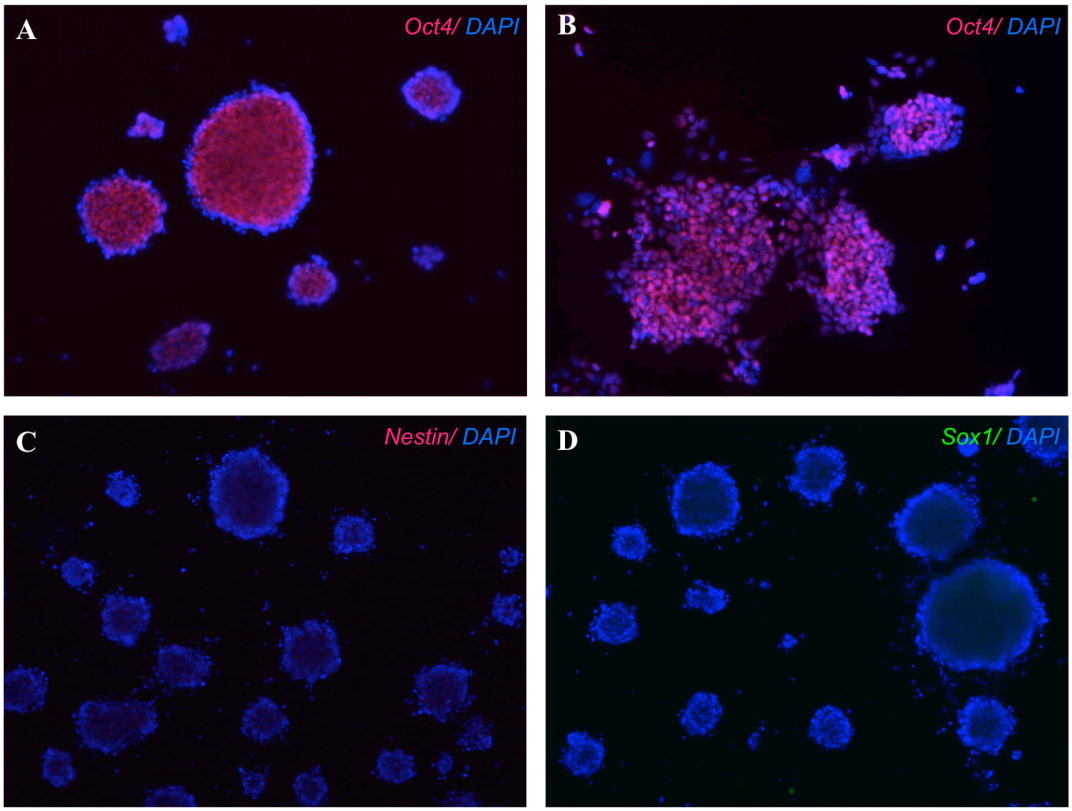


Figure 3-3: Immunostaining of mESCs grown in two alternate culture conditions. (A) Cells maintained in LIF- and serum-free '3i' medium and stained for OCT4 (red). (B) Cells maintained in LIF- and serum-containing expansion medium and stained for OCT4 (red). (C) mESCs grown under 3i conditions were uniformly nestin-negative (red). (D) mESCs grown under 3i conditions were uniformly SOX1-negative (green). DAPI nuclear counterstaining is shown in blue.

3.2.2 Embryonic stem cell proliferation analysis

Whilst mESCs grown under 3i conditions were able to undergo frequent passaging with no observable reduction in growth rate, an EdU assay (Click-iT® EdU Imaging Kit, Molecular Probes™; #C10086) was performed to confirm the proliferative capacity of this cell type. EdU, a thymidine analogue, is incorporated into newly synthesized DNA during S-phase of the cell cycle and may be detected through immunostaining with an anti-EdU primary antibody and fluorescent (Alexa Fluor®

594) secondary antibody. mESCs were incubated with media containing 10µM EdU for two hours before being fixated in 4% PFA and immunostained (Figure 3-4). Although the boundaries of individual cells and their nuclei were not clearly discernible due to the three-dimensional, dome-shaped nature of the mESC colonies, widespread EdU staining was observed in approximately half of all DAPI-counterstained nuclei.

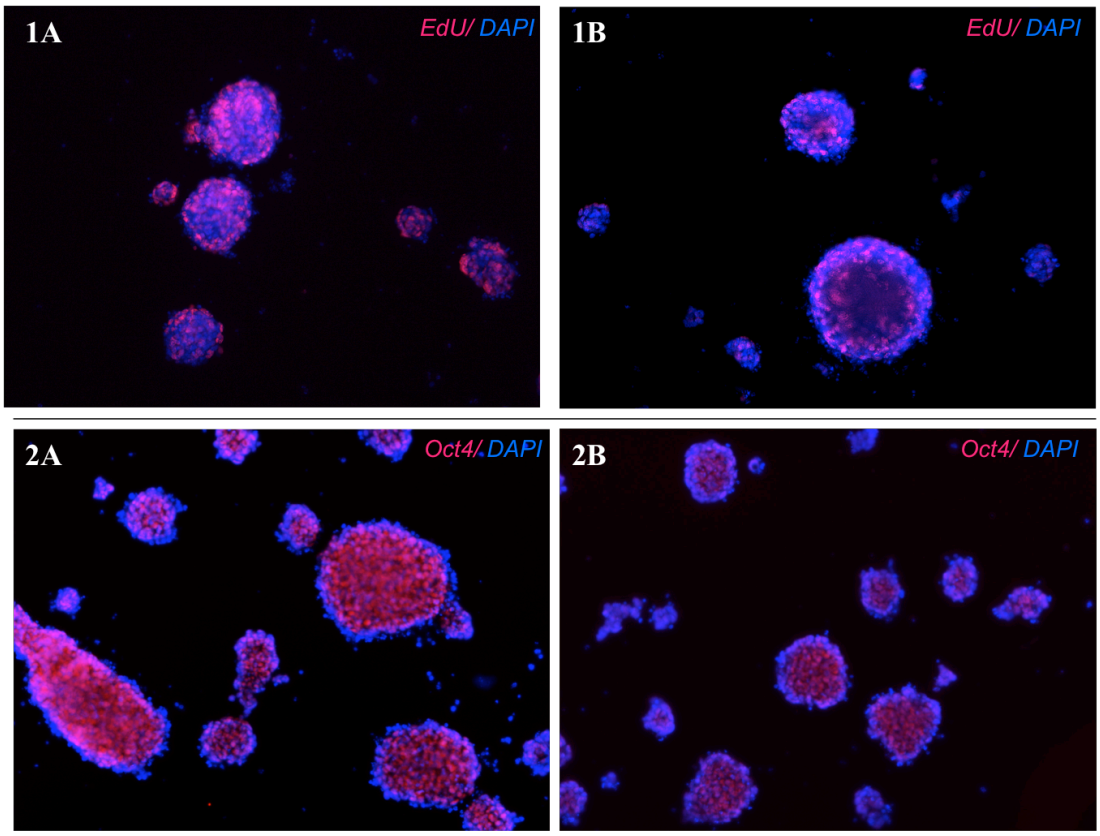


Figure 3-4: Comparison of EdU and OCT4 immunostaining of mESCs. Panel 1 shows mESCs stained for EdU (red). Panel 2 shows mESCs stained for OCT4 (red). All images show nuclei counterstained with DAPI (blue). Uniform OCT4 stained was observed across all mESC colonies, whist approximately 50% of mESCs were positive for EdU incorporation.

3.3 Primary neural stem cell culture and differentiation

A WT mNSC line derived from an adult C57BL/6N mouse was grown as an adherent monolayer on laminin-coated plasticware. NSC expansion medium was comprised of RHB-A®, a serum-free, chemically defined culture medium (StemCells, Inc.; #SCS-SF-NB-01), in combination with epidermal growth factor (EGF) and basic fibroblast growth factor (FGF-2), each at a final concentration of

10ng/ml. These cells showed a doubling time of approximately 24 hours and could be passaged extensively without any apparent change in morphology or growth rate.

Cell phenotype was confirmed by immunofluorescence for a combination of neural markers. Individual wells of 4-well culture dishes, seeded with P7 mNSCs, were stained with either anti-nestin (#ab6142) or anti-GFAP (#G9269) antibodies (Figure 3-5, 1A-B). Cells were immunoreactive for nestin, a distinguishing marker of NSCs, and GFAP-negative (Wiese et al., 2004).

Withdrawal of growth factors from the RHB-A media and addition of 10% foetal bovine serum (Gibco®; #10500-064) triggered the differentiation of mNSCs down the glial lineage to produce astrocytes, resulting in flat, stellate-like cells after just 24 hours. Cells were kept in this differentiation medium, without re-plating or passaging, for two weeks before harvesting the cells for DNA extraction. Immunostaining conducted in parallel revealed 100% of the cells were GFAP-positive after this two-week exposure to serum (Figure 3-5, 2A-B). In addition, no nestin immunoreactivity was detected above background signal, suggesting that a homogeneous population of astrocytes had been achieved under these culture conditions.

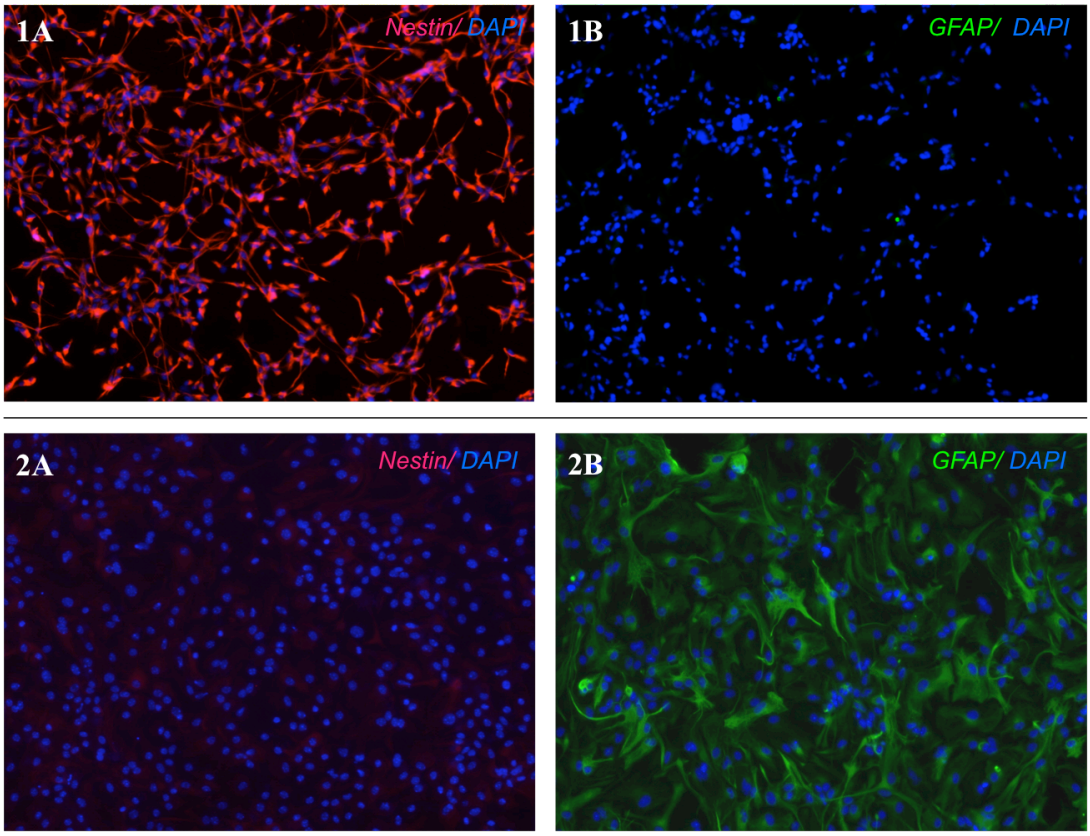


Figure 3-5: ICC staining of primary mNSCs and differentiated astrocytes. Panel 1: mNSCs under proliferating conditions showed uniform nestin expression (1A) without any visible GFAP expression (1B). Panel 2: After 14 days of differentiation through exposure to serum, mNSCs had adopted typical astrocytic morphology and were broadly nestin-negative (2A) and GFAP-positive (2B).

3.3.1 Neural stem cell and astrocyte proliferation analysis

The proliferative capacity of mNSCs was monitored using an automated live cell imaging system (IncuCyte™, Essen Biosciences) that took cell density measurements every 2 hours over a one-week period, resulting in an exponential growth pattern after an initial phase of minimal cell growth, likely due to low seeding densities (Figure 3-6A). Incubation with EdU confirmed that a high proportion of cells (~40%) were in S-phase under these expansion conditions (Figure 3-6C).

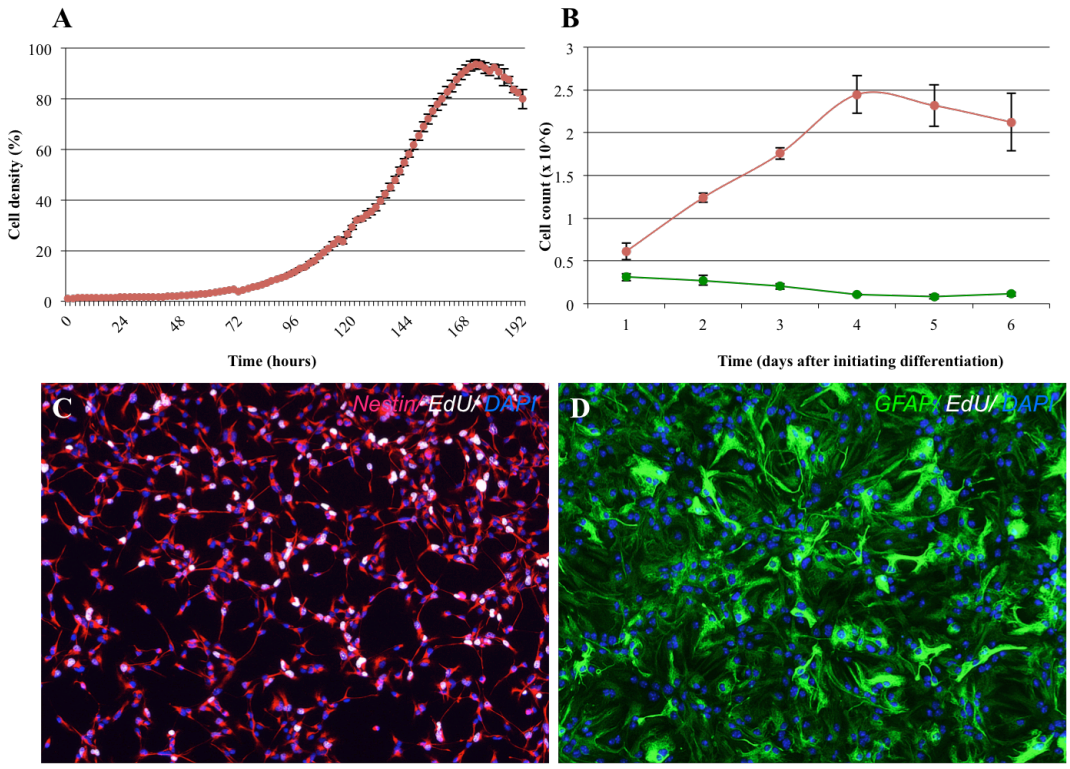


Figure 3-6: Investigating the proliferative potential of mNSCs and differentiated astrocytes. (A) Cell density measurements of primary mNSCs in expansion conditions over a one-week period show a typical sigmoid growth curve. Error bars represent the standard deviation between 3 replicate measurements. (B) Total cell counts of mNSCs grown in either expansion conditions (red) or differentiation conditions (green) over a 6-day period. Error bars represent the standard deviation between 3 replicate measurements. (C) Approximately 40% of all nuclei stained positive for EdU, indicating active proliferation in the mNSC culture. Nuclear EdU staining is shown in white; nestin staining is shown in red. (D) No nuclear EdU staining was visible in the astrocyte culture; GFAP staining is shown in green. DAPI nuclear counterstaining is shown in blue (both images).

In contrast, cell division was strongly repressed in differentiating mNSCs upon switching growth factors for serum in the culture medium. Serum is known to be a potent inhibitor of proliferation and was used at high levels (10%) for astrocytic differentiation. The effect of serum on cell proliferation was assessed by counting total cell numbers for six consecutive days after switching from expansion medium to serum-containing differentiation medium. Individual wells of 6 x 6-well plates were seeded with approximately 0.3 x 10^6 mNSCs in either expansion or serum-containing medium (3 wells per condition). After 24 hours, cells were collected from each well of the first plate and counted using an automated cell viability analyser (Vi-CELL® XR, Beckman Coulter) that excludes dead cells using trypan blue staining. These measurements were repeated every 24 hours for a total of 6 days and mean cell counts calculated for each culture condition (Figure 3-6B). Cells exposed

to expansion medium displayed fairly linear cell growth until day 4, after which a slight decrease in cell count was observed for the remaining 2 days due to overcrowding of the culture dish. Cells exposed to serum, however, exhibited complete absence of growth from day 1. Some cell death was observed under these differentiation conditions as cell numbers decreased slightly over the 6-day culture period. In addition, an EDU assay performed on astrocyte cultures 2 weeks after initiating differentiation showed no visible incorporation of EdU (Figure 3-6D), confirming the fully differentiated status of these cells.

3.4 Neural differentiation of embryonic stem cells

A knock-in *Sox1-GFP* reporter mESC line, originally derived from E14Tg2a.IV cells (Ying et al., 2003), was used in place of WT mESCs in order to facilitate the monitoring of the neural differentiation process.

Neural precursor cells were successfully generated from mESCs by replating at 1 x 10^4 cells/cm^2 in RHB-A media. RHB-A has been shown to achieve more efficient neural commitment of mESCs than N2B27, the conventional serum-free medium described by Ying and Smith (2003) in the first demonstration of neural induction in a feeder-free, adherent monolayer mESC culture (Abranches et al., 2009; Diogo et al., 2008). Differentiating mESCs were monitored for the appearance of SOX1-positive neural precursors (Figure 3-7A-F). *Sox1-GFP* expression was first visible by day 7 alongside nestin (Figure 3-7G-H), whilst a characteristic neural rosette structure was apparent by day 10 (Figure 3-8), in parallel with widespread *Sox1-GFP* expression.

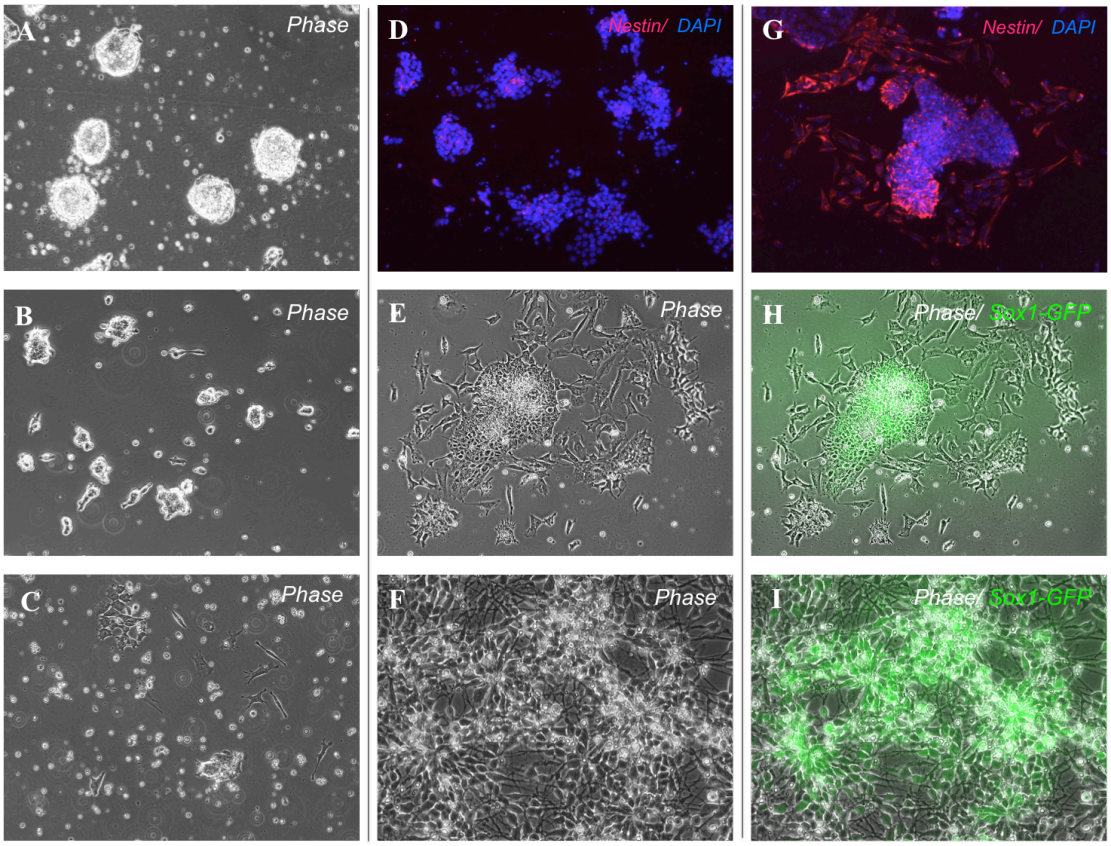


Figure 3-7: mESCs undergoing neural induction over 10 days. (A) Live imaging at day 0; mESCs under 3i culture conditions form typical spherical colonies. (B) Live imaging at day 3 of neural induction, after replating mESCs in RHB-A media. (C) Live imaging at day 5 of neural induction; significant amounts of cell death was visible at this stage. (D) ICC staining shows that nestin expression is minimal at day 6 but becomes widespread by day 7 (G). (E) Live imaging at day 7 of neural induction; elongated, bipolar cells appeared to be proliferating rapidly, outgrowing a subpopulation of flattened cells. *Sox1-GFP* expression is apparent (H) but levels are not significantly above those of background fluorescence; neural precursor cells are not yet arranged as rosette structures. (F) Live cell imaging at day 10 of neural induction; neural rosettes began to form within the culture alongside strong and widespread *Sox1* expression (I) localized to cells within neural rosettes.

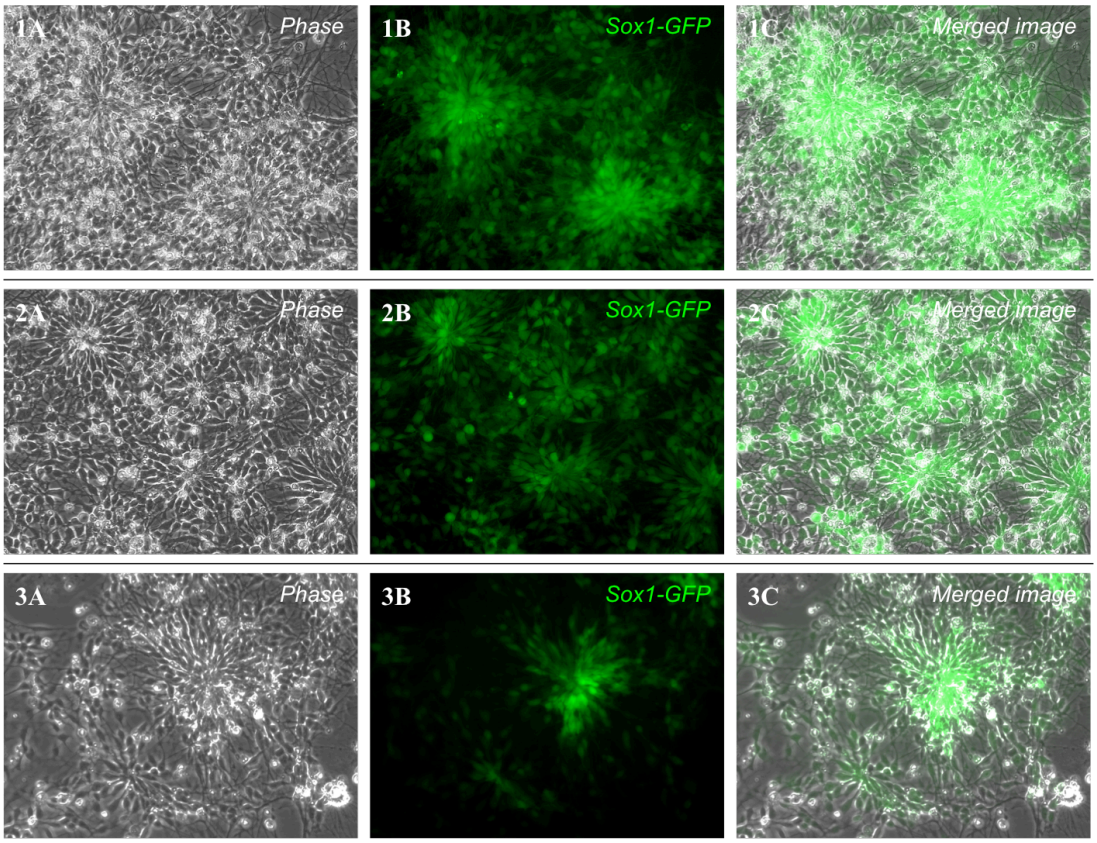


Figure 3-8: Live cell imaging at Day 10 of mESC differentiation. Phase contrast imaging (left panel) shows the formation of neural rosette structures within the culture dish. Fluorescence imaging (central panel) shows strong endogenous GFP expression concentrated within the neural rosettes.

Differentiating mESCs cultures were dissociated at day 10, re-suspended in fresh RHB-A medium containing EGF and FGF-2 and replated onto laminin-coated plasticware. Some initial cell death and cellular heterogeneity was seen upon this re-plating step, including cells with a more flattened, multipolar non-neural morphology. However, an adherent, SOX1-negative cell population rapidly outgrew these differentiated cells and subsequent passaging resulted in uniform cultures of elongated bipolar cells. These cells formed a lattice structure in the culture dish at higher densities, typical of neural stem cells (Figure 3-9B), and were maintained in culture through multiple (>8) passages without any loss in proliferation capacity or morphological change. ICC analysis confirmed these cells to be uniformly nestin-positive and GFAP-negative, in agreement with a neural stem cell identity (Figure 3-9C and D).

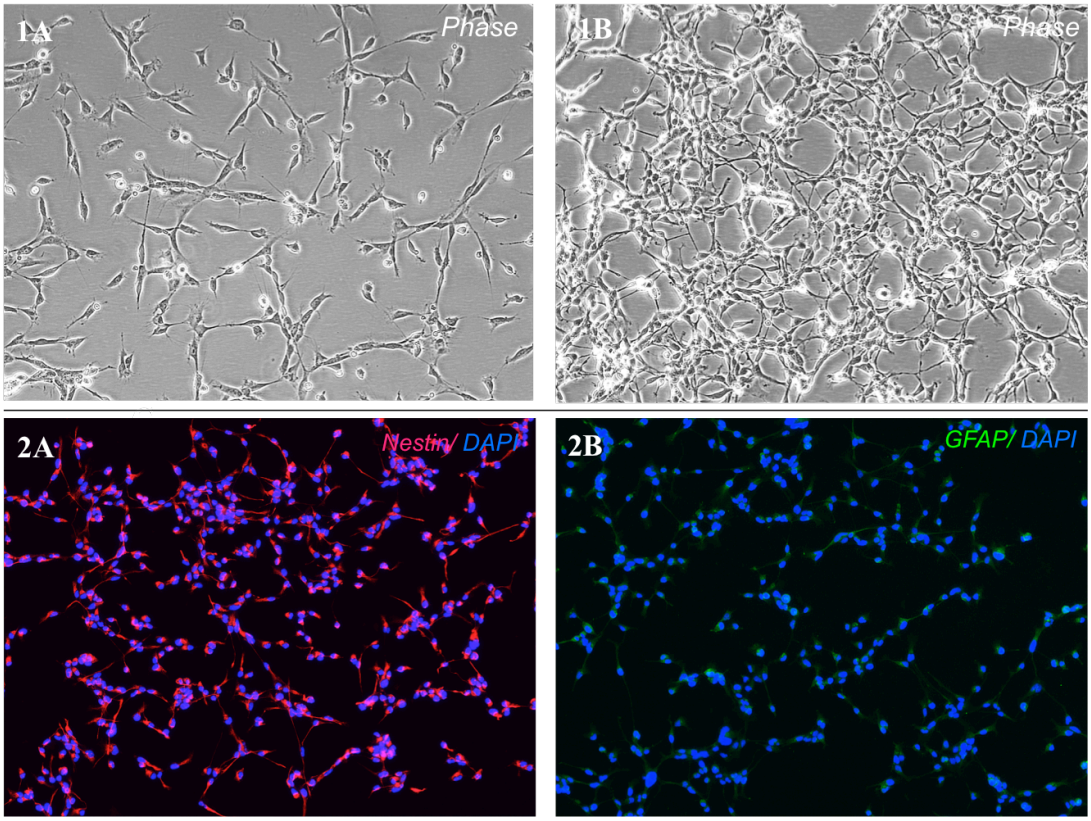


Figure 3-9: ESC-derived mNSCs in expansion medium. (1A) Live cell imaging of 'day 10' differentiated mESCs replated in neural stem cell expansion medium. (1B) ESC-derived mNSCs after three passages in expansion medium. (2A) ESC-derived mNSCs showed uniform nestin immunoreactivity (red). (2B) mNSCs were negative for GFAP immunoreactivity (green). DAPI nuclear counterstaining is shown (blue) in panel 2.

Astrocyte differentiation of embryonic-derived mNSCs was induced using the same culture conditions described previously for primary cells, through withdrawal of EGF and FGF-2 and addition to 10% serum to the neural culture medium (RHB-A). Cells were maintained under these conditions for 2 weeks before either ICC staining or cell harvesting for DNA extraction (Figure 3-10).

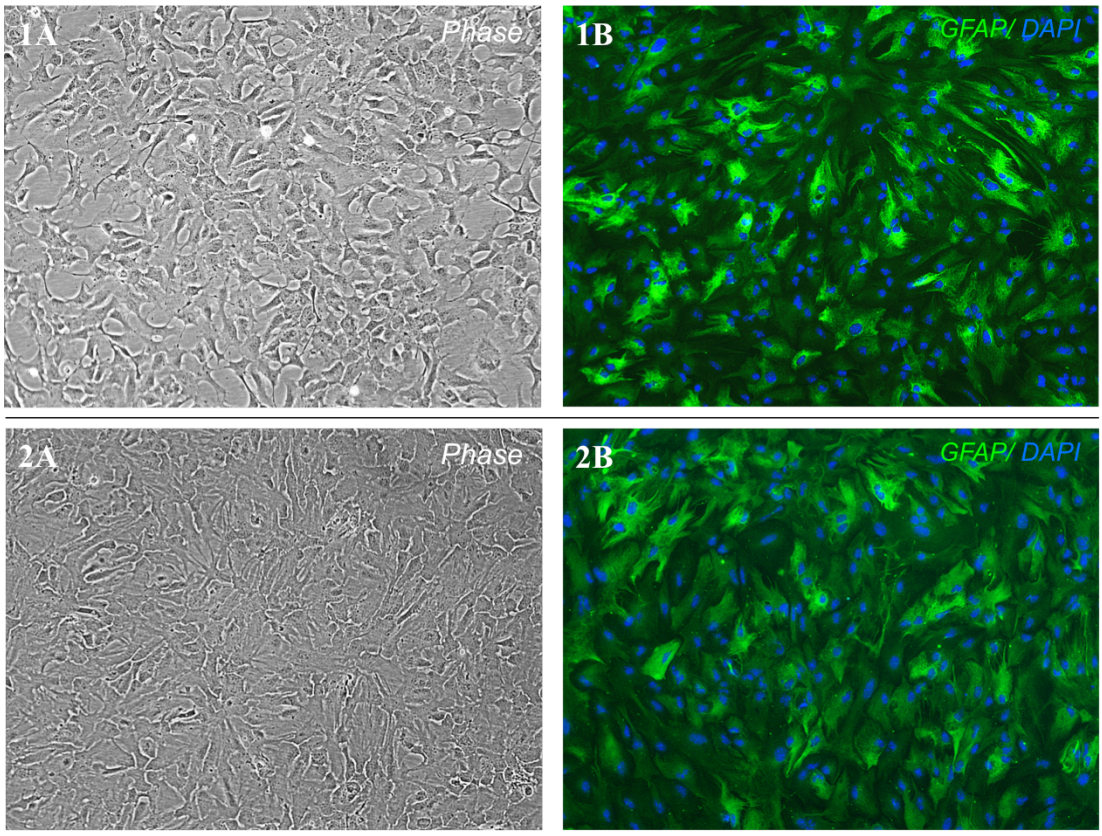


Figure 3-10: Differentiation of embryonic-derived mNSCs to astrocytes. (1A) After only 24 hours of differentiation using 10% serum-containing media, a distinctive glial morphology is observed. (1B) ICC analysis after differentiation for 24 hours shows cells broadly expressing the astrocytic cell marker GFAP (green). (2A) After 2 weeks of serum exposure, no significant differences in cellular morphology are observed. (2B) Cells remain uniformly GFAP-positive (green).

3.5 Neuronal differentiation of mNSCs

Whilst the focus of this project was on the 5mC and 5hmC dynamics associated with astrocytic fate determination, comparison of global 5hmC levels in astrocytes to other neural lineages was also desirable as it was unclear whether the significant 5hmC content associated with CNS tissue was restricted to a particular neural subtype. In an attempt to generate mature neurons *in vitro*, the *Sox1-GFP* mESC line was used to assess three variations of a neuronal differentiation protocol (Figure 3-11). In the first method (A), mESC colonies growing under 3i conditions are dissociated, re-suspended in fresh RHB-A medium and seeded onto a gelatin-coated surface at a density of 0.5-1.5 x 10^4 cells/cm^2. The cells are then left to undergo continuous differentiation in these conditions without further re-plating or changes to media composition (Ying and Smith, 2003). This protocol generated a high percentage of neuronal-like cells with thin, elongated processes extending throughout

the culture dish (Figure 3-12A-D), although cells with glial morphology were also observed in low numbers throughout the culture. These cells were maintained under differentiation conditions until day 21, when they were harvested for DNA extraction. However, cells at this stage of differentiation did not survive replating for ICC analysis, so the phenotype of these neuronal cells could not be accurately assessed.

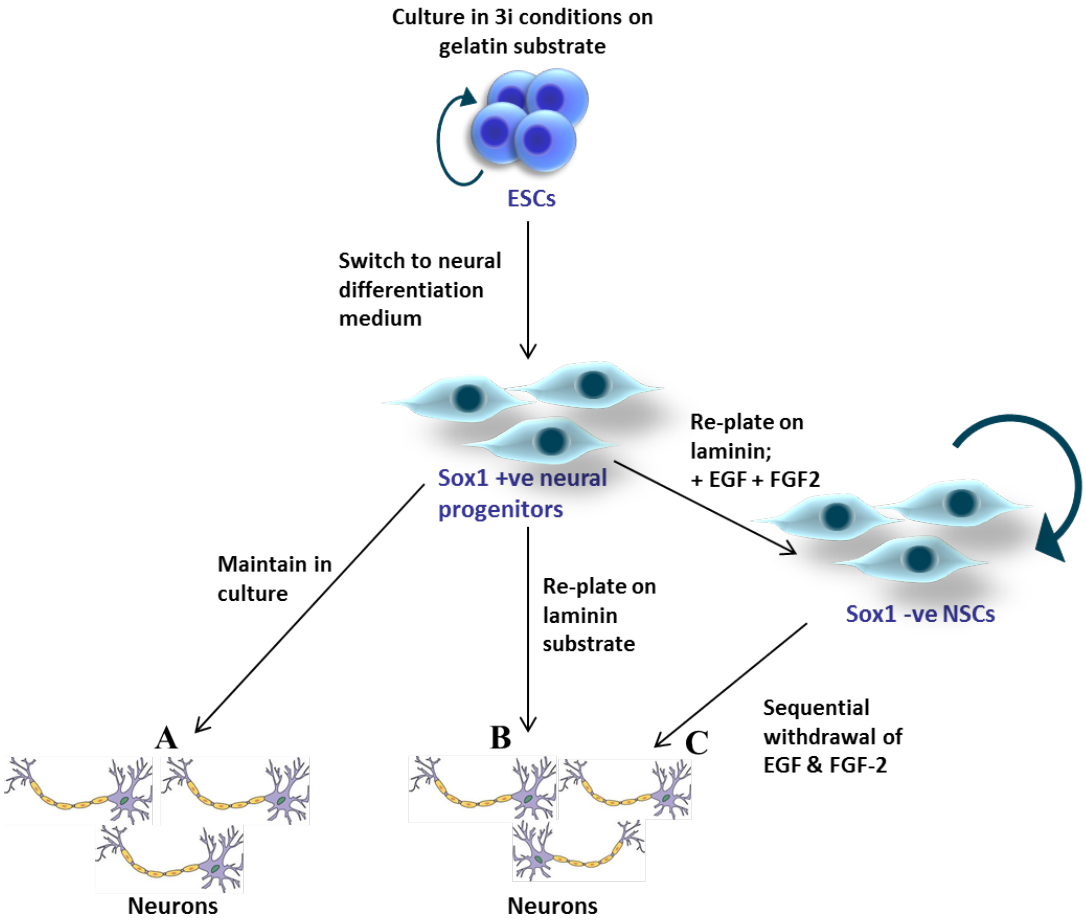


Figure 3-11: Neuronal differentiation of mNSCs. An outline of three culture methods attempted for the *in vitro* generation of differentiated neurons. SOX1-positive neural precursors were generated from mESCs as described in section 3.4. Methods (A) and (B) are based on the protocol from Ying and Smith, 2003. Method (C) is based on the protocol from Conti et al. 2005.

In a variation of the above protocol, cells may undergo a replating step at the neural precursor stage to promote a higher proportion of neuronal differentiation (method B). When SOX1-positive neural precursors are detected amongst the differentiating cell population (around day 7-10), they are dissociated and re-plated at slightly lower density (4-5 x 10^4 cells/cm^2) in RHB-A medium on a laminin substrate (Lorincz,

2006). The use of laminin over gelatin promotes increased cell attachment and re-plating is thought to promote a greater proportion of neuronal differentiation by disrupting cell-to-cell contacts and reducing cell density in the culture dish, thereby reducing the effects of autocrine factors released from neighbouring cells that promote an undifferentiated state. However, few cells successfully reattached upon replating on day 10 and did not display the characteristic morphology of neural precursors, instead appearing as a heterogeneous and somewhat differentiated population (Figure 3-12E-F). No further visible neuronal differentiation took place and cells failed to survive in culture past a few days.

The final method (C) relies on the generation of intermediary pluripotent progenitors prior to full neuronal differentiation. A stably renewing mNSC population is first established from the initial mESC culture, as described in section 3.4, before sequentially withdrawing growth factors to induce a neuronal cell fate, resulting in a final cell population containing 30 - 40% MAP2- and NeuN-positive neuronal cells (Conti et al., 2005). In this protocol, stably renewing mNSCs are dissociated, resuspended in RHB-A media containing FGF-2 only and replated at a density of 2-3 x 10^4 cells/cm^2. After 7 days of differentiation under these conditions, FGF-2 is withdrawn from the media and cells undergo a further 7 days of differentiation to induce a neuronal cell fate. Although this method did result in neuronal-like cells with long processes, cell yields were extremely low due to widespread cell death, making this method unsuitable for the generation of sufficient quantities of neurons necessary for downstream DNA methylation analysis (Figure 3-13).

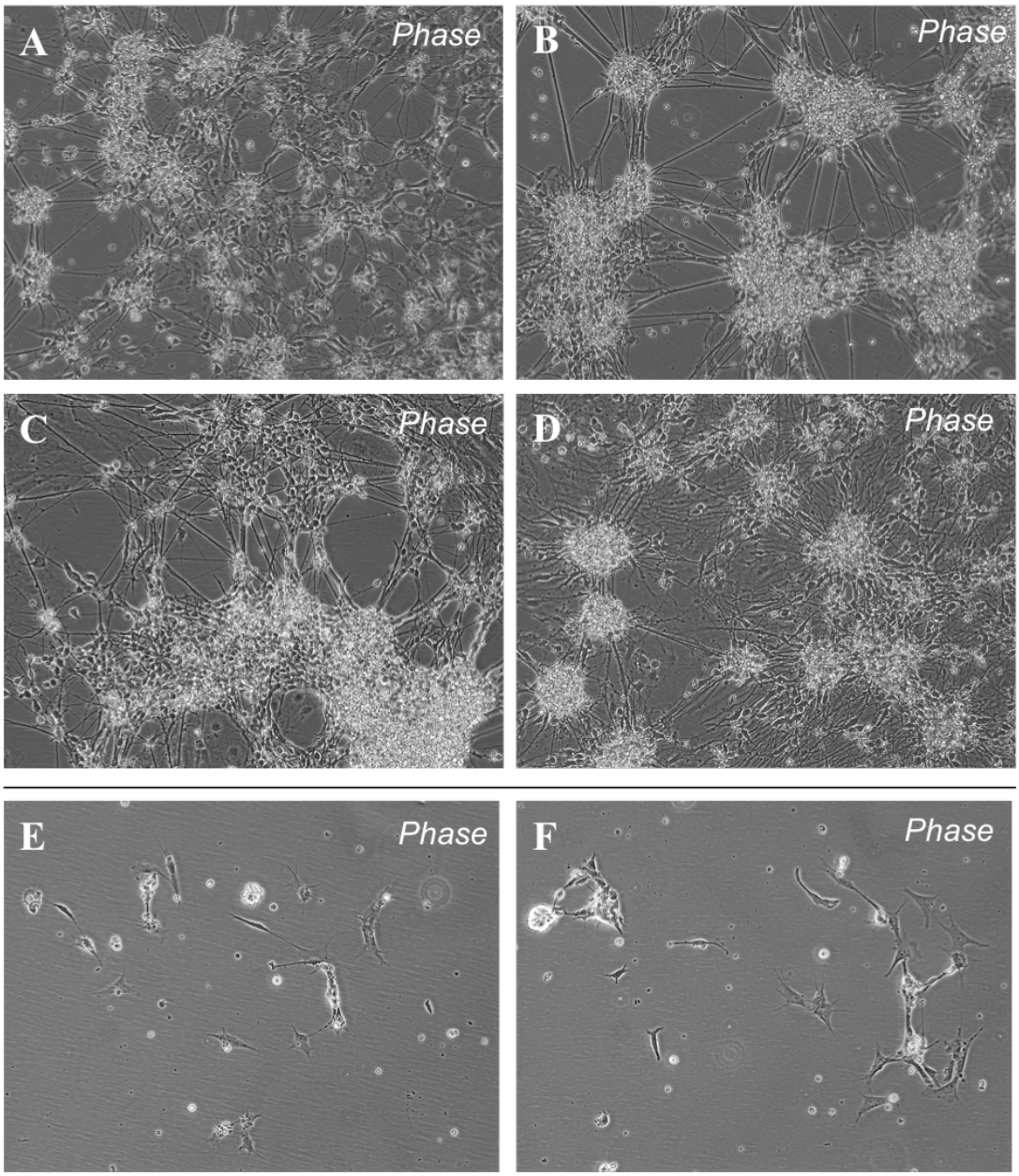


Figure 3-12: Continuous neural differentiation of mESCs as an adherent monolayer in RHB-A with and without a replating step. (A-B) Day 14 of continuous neural induction on a gelatin substrate; cells grew long processes and started to form a lattice structure within the culture dish, suggesting the onset of a neuronal cell identity. (C-D) Day 17 of continuous neural induction on a gelatin substrate. Neuronal morphology had become more widespread; some flattened non-neuronal cells remained in the culture dish. (E-F) Day 13 of neural induction after replating neural progenitors onto laminin on day 10. Very few cells survived the replating step and displayed heterogeneous morphology upon attachment.

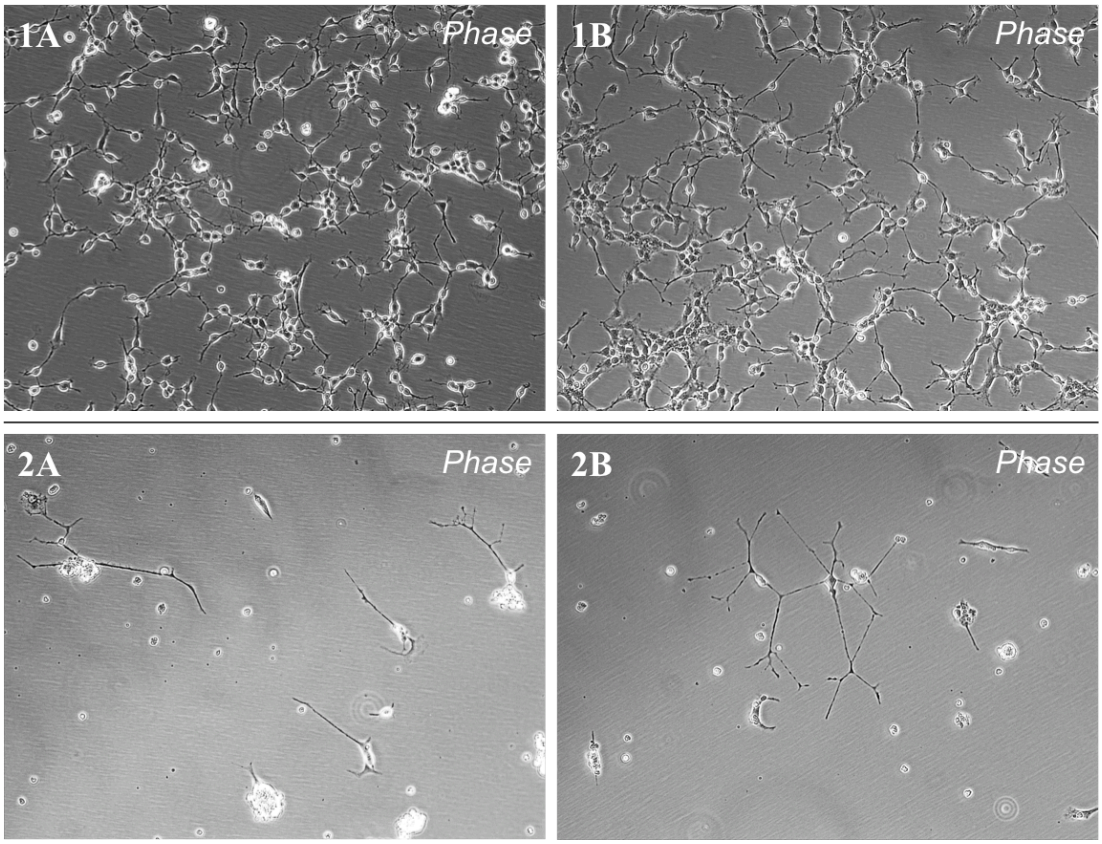


Figure 3-13: Live cell imaging of an embryonic-derived mNSC population undergoing neuronal differentiation through sequential growth factor withdrawal. Self-renewing mNSCs were replated in RHB-A with FGF-2 (10ng/ml) only to initiate differentiation. (1A) Day 4 of EGF withdrawal. (1B) Day 7 of EGF withdrawal. FGF-2 was withdrawn on this day and differentiation continued in RHB-A medium only. (2A & 2B) Day 14 of EGF withdrawal/Day 7 of FGF-2 withdrawal. Significant cell death was observed upon FGF-2 withdrawal, with only a few adherent cells remaining at the end of the 14-day differentiation period.

3.6 Discussion

An *in vitro* culture model was successfully developed and established for the generation of three defined cell populations: mESCs, mNSCs and astrocytes. Phenotypes were confirmed using a combination of morphological observation, ICC staining and proliferation analysis, performed in parallel to the expansion and/or differentiation of the corresponding cells for collection (Table 3-1).

Cell samples for methylome and hydroxymethylome analysis were all derived from the WT C57BL/6N mouse strain. Both stem cell populations (mESCs and primary mNSCs) were expanded separately *in vitro* and harvested for DNA extraction. Astrocytes were differentiated directly from primary mNSCs and harvested after a two-week culture in 10% serum-containing medium to ensure full differentiation.

10% serum exposure for a minimum of 48 hours is a traditional method for the *in vitro* production of astrocytes and is sufficient for the induction of two characteristic astrocyte markers, GFAP and glutamine synthetase, concomitant with the downregulation of nestin (Brunet et al., 2010; Loo et al., 1995; Sakai et al., 1990), although mature, morphologically homogeneous cultures of GFAP-expressing astrocytes have been previously derived in culture after 48 hours in the presence of just 1% serum (Conti et al., 2005).

In addition, the neural differentiation of *Sox1-GFP* mESCs in adherent monolayer conditions permitted the generation of embryonic-derived mNSCs, astrocytes and neuronal-like cells. These alternatively derived cell populations were used for validation of the genome-wide 5mC and 5hmC profiling data in order to assess whether observed DNA methylation changes between mESCs and primary neural cells could be recapitulated under *in vitro* differentiation culture conditions.

Cell Type	MeDIP/hMeDIP-seq analysis	oxBS-pyroseq analysis	LC-MS analysis
WT C57BL/6N mESCs	✓	✗	✓
WT C57BL/6N adult mNSCs	✓	✗	✓
WT C57BL/6N astrocytes	✓	✗	✓
E14Tg2a.IV *Sox1-GFP* mESCs	✗	✓	✓
E14Tg2a.IV *Sox1-GFP* mNSCs	✗	✓	✓
E14Tg2a.IV *Sox1-GFP* astrocytes	✗	✓	✓

Table 3-1: Summary of all cell culture samples collected for subsequent DNA analysis.

Chapter 4 Methylome and hydroxymethylome analysis of neural differentiation

Work carried out in collaboration:

MeDIP- and hMeDIP-seq DNA libraries were sequenced on an Illumina HiSeq 2000 at Illumina Cambridge Ltd, UK. Bioinformatics analysis of raw sequencing data using the MEDIPS pipeline was carried out by Dr Gareth Wilson (UCL Cancer Institute). LC-MS analyses of genomic DNA samples were performed by Martin Bachman (Department of Chemistry, University of Cambridge).

4.1 Introduction

DNA methylation at the 5' position of cytosine is a well-characterised epigenetic modification in mammalian DNA and is known to play an important role in a diverse range of cellular processes, including genomic imprinting, X-chromosome inactivation, retrotransposon gene silencing and tissue-specific gene regulation (Bird, 2002; Reik, 2007). 5-methylcytosine (5mC) is thought to be a heritable, stable epigenetic modification that nevertheless undergoes dynamic changes during normal development in both somatic and germ line tissues (Hajkova et al., 2002; Meissner et al., 2008).

DNA methylation patterns are established and maintained by the DNA methyltransferase (DNMT) family of enzymes. Knockout studies in mice targeting the DNMT gene family have shown that homozygous deletion of either the maintenance methyltransferase DNMT1 or the *de novo* methyltransferases DNMT3A/B results in embryonic or early postnatal lethality, suggesting a critical functional role for 5mC in normal development (Li et al., 1992; Okano et al., 1999). DNMT3A and DNMT3B display distinct expression patterns in the developing central nervous system (CNS), with DNMT3B highly expressed during early neurogenesis (E11-E15), whilst DNMT3A levels are initially low but increase during early postnatal development within postmitotic neurons before being downregulated in the mature adult CNS (Feng et al., 2005).

5mC-mediated transcriptional regulation has been widely implicated in CNS development and mutations in methyl-CpG binding domain (MBD) proteins have been linked to neurological disorders such as Rett Syndrome (Goto et al., 1994; Guy et al., 2001). There is also increasing evidence supporting a role for 5mC in the dynamic regulation of gene expression patterns underlying neural plasticity (Feng et al., 2007; Hsieh and Gage, 2004). Neuronal activity has been shown to trigger changes in DNA methylation, MBD protein binding affinity and gene transcription levels, suggesting that 5mC-mediated chromatin remodelling is critical for normal neuronal gene regulation (Chen et al., 2003; Martinowich et al., 2003).

The recent discovery of 5-hydroxymethylcytosine (5hmC), an oxidation product of 5mC catalysed by the ten-eleven-translocation (TET) hydroxylase enzymes, as a novel constituent of mammalian DNA has opened up new avenues of research into the role of DNA methylation in health and disease (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). There is particular interest surrounding the potential function of 5hmC in CNS development due to the unusually high levels of 5hmC recorded in both human and mouse brain compared to other somatic tissues, immortalised cell lines and stem cells (Globisch et al., 2010; Munzel et al., 2010; Song et al., 2011). However, few methods were available at the time for the accurate quantification and profiling of this novel cytosine modification, although two methods (oxBS-seq and TAB-seq) have subsequently been developed for single-base-resolution 5hmC sequencing (Booth et al., 2012; Yu et al., 2012). Sodium bisulphite conversion coupled to high-throughput sequencing (BS-seq) is considered the gold standard for DNA methylation profiling and allows single base resolution of 5mC (Lister et al., 2009) but cannot discriminate between 5mC and 5hmC, leading to potential overestimation of the 5mC content of DNA (Huang et al., 2010; Nestor et al., 2010). Several enrichment-based techniques quickly emerged for genome-wide 5hmC profiling, including hMeDIP (Ficz et al., 2011; Stroud et al., 2011; Williams et al., 2011; Wu et al., 2011), anti-CMS and GLIB (Pastor et al., 2011), all of which can be used to immunoprecipitate 5hmC-containing DNA fragments for sequence analysis using arrays or high-throughput sequencing.

In order to explore the 5mC/5hmC changes that occur along astrocytic lineage differentiation, genome-wide 5mC and 5hmC analysis was performed for three time points at defined cellular stages: mESCs, primary mNSCs and *in vitro* differentiated astrocytes. MeDIP-seq is a well established genome-wide profiling method offering moderate resolution (100-300bp) at significantly lower cost than single-base resolution, whole-genome profiling methods such as BS-seq (Down et al., 2008; Weber et al., 2005). MeDIP-seq may also be automated for higher throughput and increased reproducibility between samples (Butcher and Beck, 2010). Automated MeDIP- and hMeDIP-seq assays were therefore selected for methylation profiling on the basis of reliability and specificity towards either 5mC or 5hmC (depending on the antibody used), whilst enabling ease of comparison between the final datasets due to similar experimental protocols.

4.2 Genome-wide DNA methylation profiling of neural differentiation

Genome-wide 5mC and 5hmC profiles were generated for three distinct cell types (mESCs, mNSCs and astrocytes) in triplicate using a combined MeDIP- and hMeDIP-seq approach, resulting in a total of nine methylomes and nine hydroxymethylomes. MeDIP- and hMeDIP-seq DNA libraries were sequenced on an Illumina HiSeq 2000 (100bp paired-end reads) and the resulting fastq data files were processed using the MeDUSA analysis pipeline (Wilson et al., 2012). Briefly, raw sequencing reads were aligned to the mm10 reference genome using BWA (Li and Durbin, 2009), filtered for redundant and unpaired reads using SAMtools (Li et al., 2009) and normalised for fragment length, resulting in final read counts in the range of 9.5 to 20 million uniquely-mapped paired-end reads per DNA library (Table 4-1, column 2). The final percentages of uniquely mapped reads were lower than expected, likely due to the existence of PCR duplicates arising during library amplification. Individual methylation and hydroxymethylation peaks were defined using MACS (Zhang et al., 2008).

The correlation between MeDIP- and hMeDIP-seq replicates was calculated to provide an estimate of concordance in genome-wide sequence coverage within each biological sample (Supplementary Table 2). The correlations were very high between replicates for all three samples (mESCs, mNSCs and astrocytes) and both types of

DNA sequencing library (MeDIP and hMeDIP). For hMeDIP-seq libraries, correlations were all > 0.95, whilst MeDIP-seq correlations were slightly lower (> 0.86).

Sample type	High quality, aligned read count (mean)	Final non-redundant read count (mean)	Final alignment rate (%)
MeDIP-seq			
Astrocytes	39,617,891	13,109,040	33.1%
Neural stem cells	42,833,593	9,487,916	22.2%
Embryonic stem cells	56,426,103	17,781,913	31.5%
hMeDIP-seq			
Astrocytes	40,290,816	18,142,576	45.0%
Neural stem cells	33,902,991	19,982,046	58.9%
Embryonic stem cells	33,062,430	17,798,058	53.8%

Table 4-1: Sequencing read counts for MeDIP and hMeDIP libraries. Three distinct DNA libraries were prepared and sequenced for each sample type and read counts shown in the table above represent the mean value of three replicates.

Dynamic changes in 5mC and 5hmC distribution were assessed through the identification of differentially methylated and hydroxymethylated regions (DMRs and hDMRs, respectively) associated with two defined stages of neural differentiation: mESCs to mNSCs, and mNSCs to astrocytes. These regions were defined as 500bp windows of differential enrichment between triplicate methylation or hydroxymethylation peak-sets, using a stringent false discovery rate (FDR) of 5%.

Most observed changes occurred in the transition from an embryonic to neural stem cell fate, with a total of 5,925 DMRs and 321 hDMRs found between these two cell types. In contrast, just 140 DMRs and 2 hDMRs were called between mNSCs and astrocytes (Table 4-2). The majority of DMRs (90%) called between mESCs and mNSCs represented gains in methylation at this transition (hyper-DMRs). All 140 DMRs observed between mNSCs and astrocytes also represented hypermethylation, suggesting a pattern of increasing genome-wide methylation as a cell transitions from a pluripotent to terminally differentiated cell fate.

Overall, the total number of hDMRs detected between the different hydroxymethylomes was approximately 20-fold less than the number of DMRs, perhaps reflecting the significantly lower amounts of global 5hmC observed in mammalian genomic DNA compared to 5mC. If 5hmC is a true intermediate in an active demethylation mechanism, it is possible that 5hmC turnover occurs too rapidly to be captured effectively by such genome-wide profiling methods.

	mNSCs vs. mESCs	Astrocytes vs. mNSCs
5mC		
Hypo-DMRs	559	0
Hyper-DMRs	5366	140
5hmC		
Hypo-hDMRs	119	2
Hyper-hDMRs	202	0

Table 4-2: Summary of the total number of differentially methylated and hydroxymethylated regions called between sample types. The majority of DMRs and hDMRs are found between mESCs and mNSCs. Few changes in 5mC and 5hmC distribution are identified between mNSCs and astrocytes.

4.3 Feature annotation of differentially methylated and hydroxymethylated regions

Genomic feature annotation of identified DMRs and hDMRs was performed through the MeDUSA analysis pipeline, which assigned a DMR or hDMR to a particular feature if either 25% of the DMR fell within the defined feature or if at least 25% of the feature was covered by the DMR in question. For calculations of feature enrichment, a minimum 50% overlap was required in order for an association to be made between a DMR and genomic feature.

During the mESC to mNSC transition, hypermethylated regions appear biased towards genic sites, with 68% of all hyper-DMRs falling within genes, compared to just 32% located in intergenic regions (Figure 4-1A). Significant percentages of genic hyper-DMRs are associated with CpG islands, shores and promoter regions (18%, 23% and 26% respectively), but significant enrichment is observed only within exons (Bonferroni-adjusted p-value = 1.83×10^{-186}). In contrast, hypo-DMRs are distributed fairly evenly between genic and intergenic locations (57% and 43%

respectively) and less than 10% are located within promoters or CpG-dense regions. However, these hypo-DMRs are nevertheless significantly enriched in promoters (Bonferroni-adjusted p-value = 1.06×10^{-12}), CpG shores (Bonferroni-adjusted p-value = 1.04×10^{-7}) and intragenic regions (both exons and introns, Bonferroni-adjusted p-values equal 2.24×10^{-6} and 1.88×10^{-5}, respectively). During the differentiation of mNSCs to astrocytes, only hyper-DMRs are observed. Of the 45% located within genes, 42% are associated with introns and less than 4% are associated with CpG islands, shores or promoter regions. It therefore appears that DMRs located within promoters and other CpG-dense regions undergo mostly hypermethylation during the mESC to mNSC transition and few additional gains occur during subsequent differentiation to astrocytes.

The genomic attributes of DMRs were also assessed using EpiExplorer, an interactive web tool that allows annotation of custom user datasets with a wide range of genomic and epigenomic features (Halachev et al., 2012). Randomised control sets may automatically be generated for each custom dataset uploaded to EpiExplorer by reshuffling the genomic coordinates of the user-upload regions in the genome, allowing comparative analysis of both control and custom datasets to assess enrichment for particular genomic annotations (Figure 4-2).

Both hypo- and hyper-DMR sets show significant enrichment compared to controls in regions associated with regulatory features including gene promoters, H3K4me1, H3K4me3, DNA polymerase II (Pol-II) binding and DNaseI hypersensitive sites and conversely a strong depletion in repetitive DNA sequence, indicating that methylation changes between mESCs and mNSCs are selectively associated with functionally relevant regions.

With the exception of two hypo-hydroxymethylated regions, all 321 hDMRs are observed during the mESC to mNSC transition, rather than between mNSCs and astrocytes (Table 4-2). Comparison of the genomic distribution of hypo- and hyper-hDMRs reveals little difference between the two categories (Figure 4-1B), with over 50% of both hDMR categories found within intergenic regions. However, hypo-hDMRs are significantly enriched within both CpG islands and shores (Bonferroni-

adjusted p-values equal 1.38 x 10^{-5} and 3.25 x 10^{-4}, respectively). No significant enrichment for particular genomic features is observed for hyper-hDMRs.

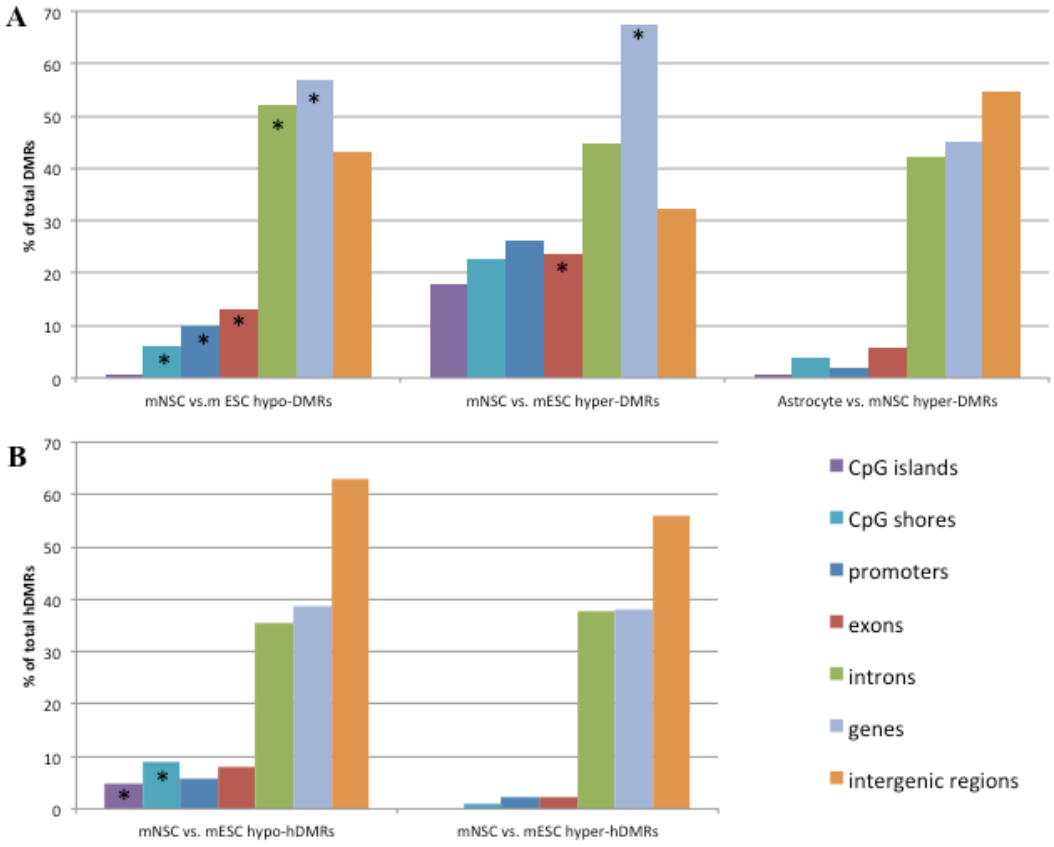


Figure 4-1: Feature annotation of all DMRs and hDMRs identified between either the mESC to mNSCs or mNSC to astrocyte transition. A) Distribution of all DMRs according to genomic feature. B) Distribution of all hDMRs according to genomic feature. Categories found to be significantly enriched (Bonferroni-adjusted p-value < 0.05) for DMRs are labelled with an asterisk.

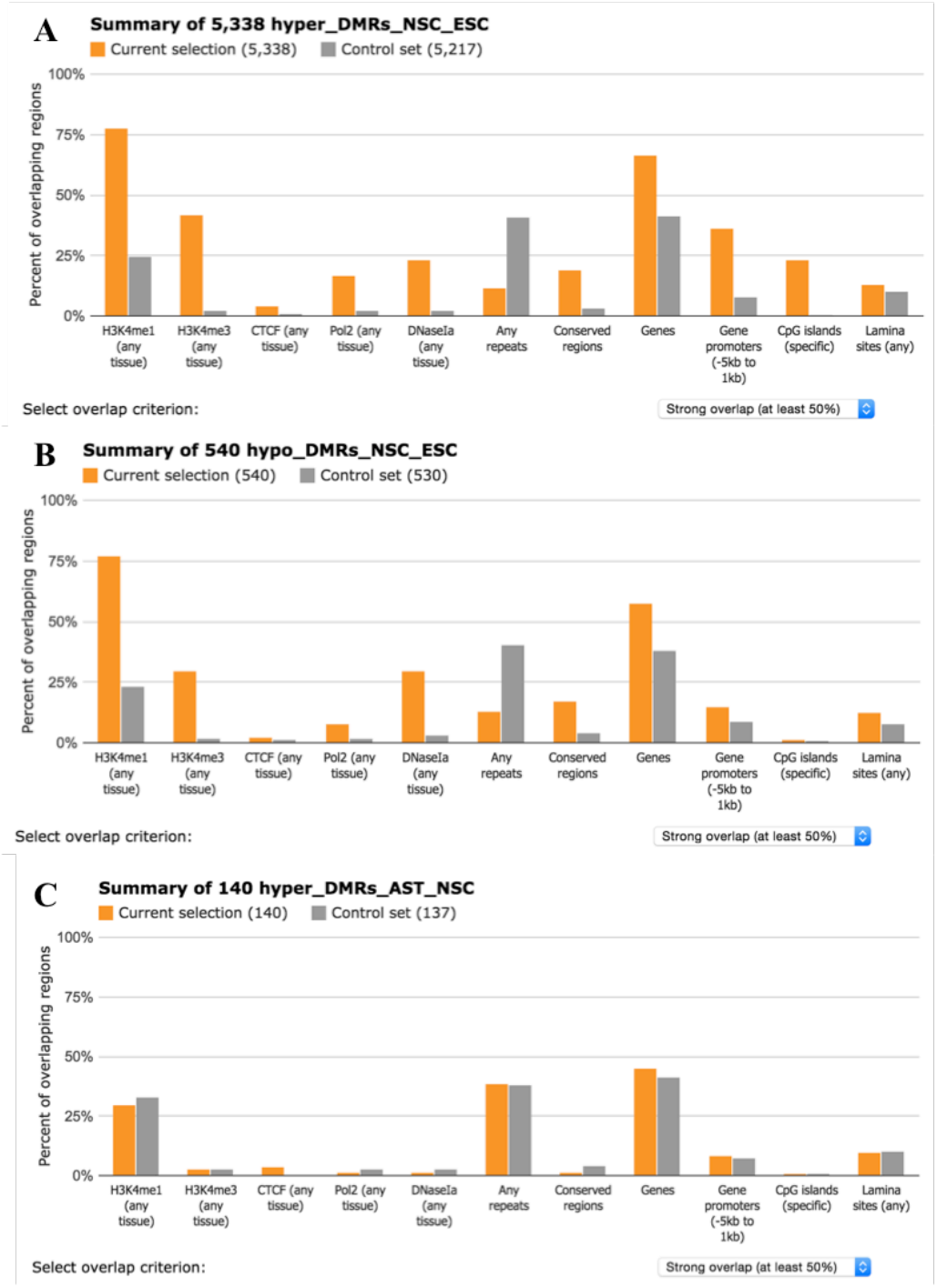


Figure 4-2: EpiExplorer analysis summaries for DMR datasets compared to randomised controls. The bar charts compare the percentage overlap of DMRs (orange bars) falling within genomic features (x-axis) to randomised control regions (grey bars). A) The percentage overlap between various genomic features and hyper-DMR regions identified between mNSCs and mESCs. B) The percentage overlap between various genomic features and hypo-DMR regions identified between mNSCs and mESCs. C) The percentage overlap between various genomic features and hyper-DMR regions identified between astrocytes and mNSCs. A DMR was associated with a particular genomic feature if they shared at least a 50% overlap in sequence coverage. Note that in some cases, the total numbers of hypo- and hyper-DMR regions examined here are slightly lower than those reported in **Table 4-2** due to a few unsuccessful conversions of DMR coordinates from the mm10 to mm9 genome assembly prior to EpiExplorer analysis.

Analysis of the DNA sequence content of the identified 5hmC hotspots revealed a bias of hDMRs towards repeat regions of the genome, unlike DMRs which are predominately located in non-repetitive sequence. Only 13% and 27.4% of all hyper- and hypo-DMRs, respectively, fall within repetitive regions of the genome (Figure 4-3A-B). In contrast, approximately 80% of both hyper- and hypo-hDMR sets are associated with repeat features (Figure 4-3C-D). High enrichment levels for repetitive DNA sequences has previously been noted for both MeDIP- and hMeDIP-seq profiles (Matarese et al., 2011), whilst similar enrichment is lacking in 5mC profiles generated using alternative methods, such as MethylCap-seq and BS-seq. This may be explained at least in part by antibody bias towards regions of high cytosine density, preferentially immunoprecipitating unmethylated but cytosine-rich DNA over methylated regions with lower cytosine density.

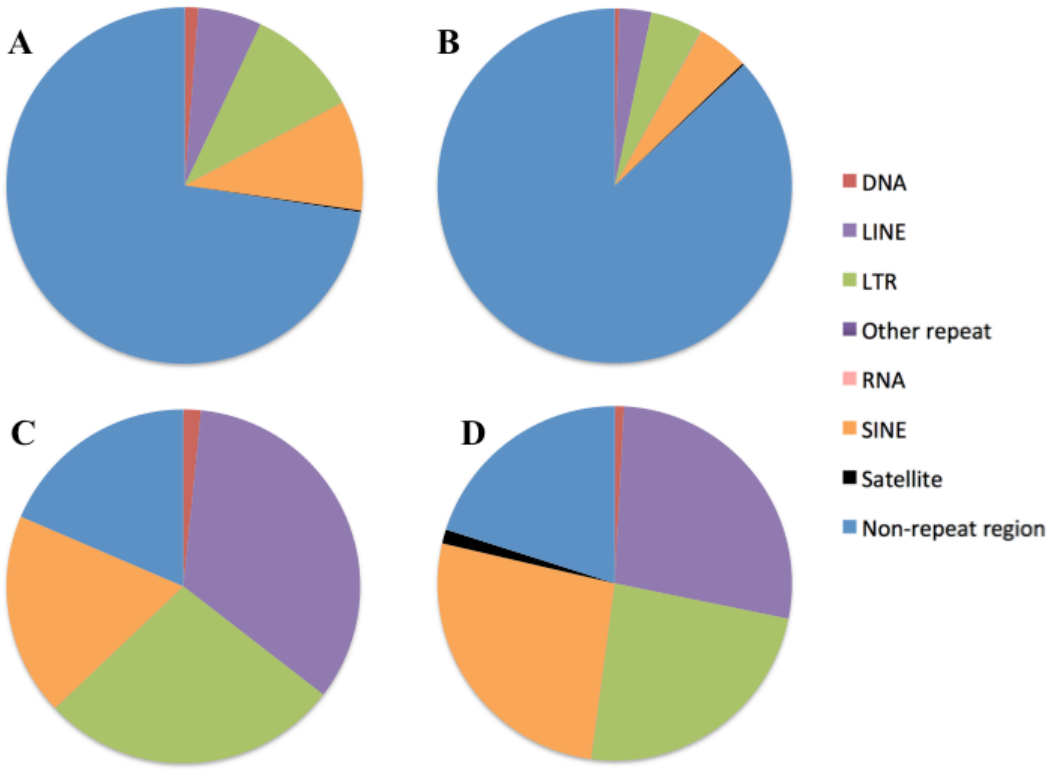


Figure 4-3: DMRs and hDMRs identified during the mESC to mNSC transition show distinct distributions within repetitive DNA sequences. Hypo-DMRs (A) and Hyper-DMRs (B) predominately fall within non-repetitive regions of the genome, shown in blue. However, the majority of hypo-hDMRs (C) and Hyper-hDMRs (D) are found within repeat regions: less than 25% of all differentially hydroxymethylated regions are associated with unique DNA sequence.

Further interrogation of the genomic attributes of hDMRs using EpiExplorer revealed that both hypo- and hyper-hDMR datasets show enrichment for repeat DNA

regions compared to their randomised controls, but not for genic regions or histone-associated sites, indicating that these hDMRs are unlikely to be functionally significant (Figure 4-4). In fact, hDMRs seem to be relatively depleted at H3K4me1 peaks, a modification frequently associated with enhancers (Heintzman et al., 2007). Nevertheless, hypo-hDMRs do appear to be enriched at gene promoters (defined as between -5kb and 1kb of the transcriptional start site) and Pol-II binding sites, which may have biological implications.

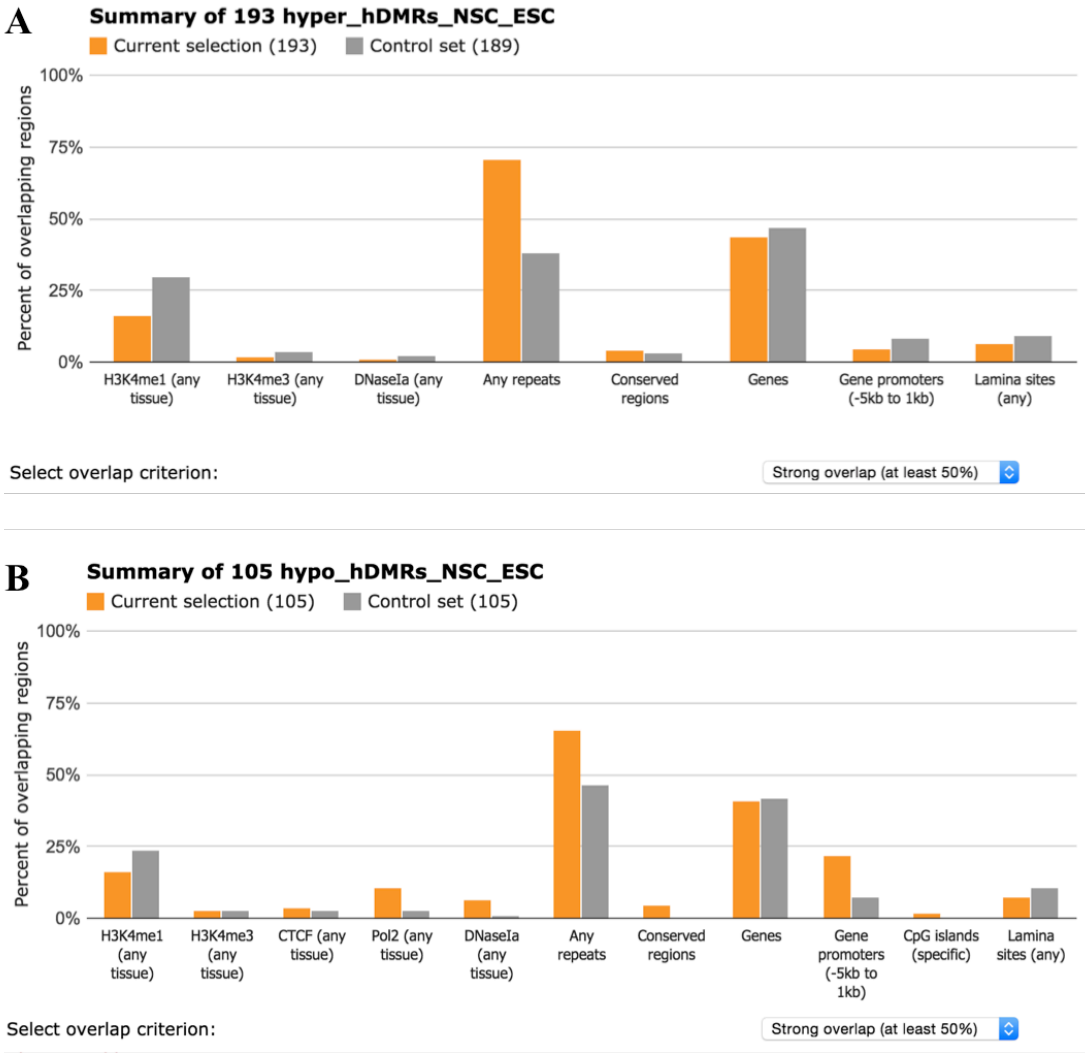


Figure 4-4: EpiExplorer analysis summaries for hDMR datasets compared to randomised control datasets. A) The percentage overlap between various genomic features (x-axis) and both hyper-hDMR regions (orange bars) and control set regions (grey bars). B) The percentage overlap between various genomic features (x-axis) and both hypo-hDMR regions (orange bars) and control set regions (grey bars). For both hDMR sets, the number of regions overlapping repeats is higher than expected by chance, suggesting enrichment for repetitive DNA. Note that the total numbers of hypo- and hyper-hDMR regions examined here are slightly lower than those reported in **Table 4-2** because the original hDMR coordinates (mm10 assembly) needed to be converted to the mm9 assembly prior to EpiExplorer analysis, resulting in a few unsuccessful lift-overs.

4.4 Functional annotation of differentially methylated regions

A combination of gene ontology analysis and individual examination of DMR regions was used in order to assess the potential functional relevance of the observed 5mC and 5hmC changes and identify those likely to hold the most biological significance during the mESC to astrocyte transition.

Differentially methylated and hydroxymethylated regions were assigned an associated gene ID if the regions fell within 100kb upstream or 50kb downstream of a known gene, according to Ensembl's annotation of the GRCm38 genome assembly (Table 4-3). The top 10 most statistically significant hyper- and hypo-DMRs called between the mESC and mNSC transition and their associated genes are detailed in Supplementary Table 3 and Supplementary Table 4, respectively.

During the mNSC to astrocyte transition, over 40-fold fewer regions of differential methylation were identified compared to the mESC to mNSC transition; all 140 regions represented hypermethylation events in astrocytes, suggesting that previous gains in methylation between mESCs and mNSCs (5366 regions in total) were conserved upon subsequent astrocytic differentiation. Similarly, comparison of the coordinates of hyper-DMRs gained during mNSC to astrocyte differentiation to those of the hypo-DMRs called between mESCs and mNSCs reveal no overlap: all initial 5mC changes were retained upon subsequent differentiation to astrocytes. Visual inspection of candidate DMRs using the UCSC genome browser (Figure 4-7 and Figure 4-8) confirmed that these methylation changes were unidirectional during the differentiation of mESCs to astrocytes.

The Database for Annotation, Visualisation and Integrated Discovery (DAVID) online functional annotation tool (v6.7) was used for further analysis of these DMR-associated gene lists. During the mESC to mNSC transition, gene-sets associated with both hypo- and hyper-DMRs showed the strongest enrichment for genes upregulated in brain tissue (Supplementary Figure 2). In addition, the second most enriched tissue expression category for the hypo-DMR gene-set was "neural stem cell", suggesting that these loss-of-methylation regions may be involved in the upregulation of genes crucial for neural stem cell determination.

	mNSCs vs. mESCs		Astrocytes vs. mNSCs	
	Total DMR number	**Associated genes**	**Total DMR number**	**Associated genes**
5mC				
Hypo-DMRs	559	430	0	0
Hyper-DMRs	5366	2906	140	107
5hmC				
Hypo-hDMRs	119	40	2	2
Hyper-hDMRs	202	95	0	0

Table 4-3: The number of annotated genes associated with all DMR and hDMR regions called between the mESC to mNSC transition (left) and the mNSC to astrocyte transition (right).

Gene ontology (GO) analysis revealed high levels of enrichment for genes associated with organismal developmental processes, cellular differentiation, nervous system development and transcriptional regulation for both DMR-associated gene sets (Supplementary Figures 3 and 4), further indicating that these methylation changes may have an important functional role in the specification of neural stem cell identity.

Functional annotation of the gene list associated with all 140 hyper-DMRs identified between the mNSC to astrocyte transition also showed significant enrichment for genes upregulated in brain, brain cortex and head tissue (Supplementary Figure 5), supporting a potential functional role for these DMRs in regulating neural specification. Genes upregulated in brain (39 total) include Cntnap2, Gata3, Atxn1, Nkx2-1, Tenm2 and Tenm4. However, GO analysis did not reveal any other terms suggestive of a specific association with neural developmental processes. Some of the most significantly enriched GO terms were those ubiquitously associated with normal eukaryotic cell function, including protein binding, cell part, metabolic processes and protein phosphatase type 2A regulator activity (Supplementary Figure 6).

Furthermore, GO analysis using the Genomic Regions Enrichment of Annotations Tool (GREAT, v3.0.0) returned similar annotation results for the identified DMR regions between mESCs and mNSCs as obtained by DAVID analysis of the DMR-associated gene lists. Hyper-DMR regions were highly enriched for genes associated

with GO biological process terms such as embryonic morphogenesis, pattern specification process, cell fate commitment and rostrocaudal neural tube patterning (Figure 4-5A). The most significantly enriched GO biological process (embryonic morphogenesis) was associated with 777 hyper-DMR regions and 233 genes, including many involved in early embryonic development such Nanog, Nodal, Pax6, Otx1 and 2, Tbx2, Notch1, En1, Zfp281 and multiple Hox genes. Enrichment was also observed for genes upregulated during the early stages of embryonic and neural development and those associated with particular mouse phenotypes such as abnormal forebrain development and other disorders of the CNS (Figure 4-5B-C).

Similarly, GREAT analysis of hypo-DMRs identified during the mESC to mNSC transitions found significant enrichment for genes associated with biological process terms including cell proliferation in the forebrain and stem cell differentiation, genes upregulated in various CNS tissues and genes associated with mouse phenotypes such as abnormal neuron differentiation and abnormal nerve morphology (Figure 4-6). GREAT analysis of the 140 hyper-DMRs identified during the mNSC to astrocyte transition did not reveal any significant enrichment for particular GO terms covering biological processes, molecular function or mouse phenotype. However, gene enrichment was found for nine tissue expression terms in total, including branchial arch, hippocampus granule cell layer and the optic nerve (Supplementary Figure 11). Although branchial arch tissue showed the most statistically significant association with 34 of the DMR regions, the hippocampus granule cell layer showed the greatest fold enrichment out of the nine GO terms (15.6), over three times higher than all other tissue categories, suggesting that some of these hyper-DMR regions may be functionally involved in astrocytic differentiation. The relative lack of enrichment overall is probably due to the small number of input DMR regions.

Taken together, the various GO analyses of DNA methylation changes between mESCs and mNSCs point to a strong association with several neural developmental processes and support a functional role for 5mC in the specification of neural cell fate.

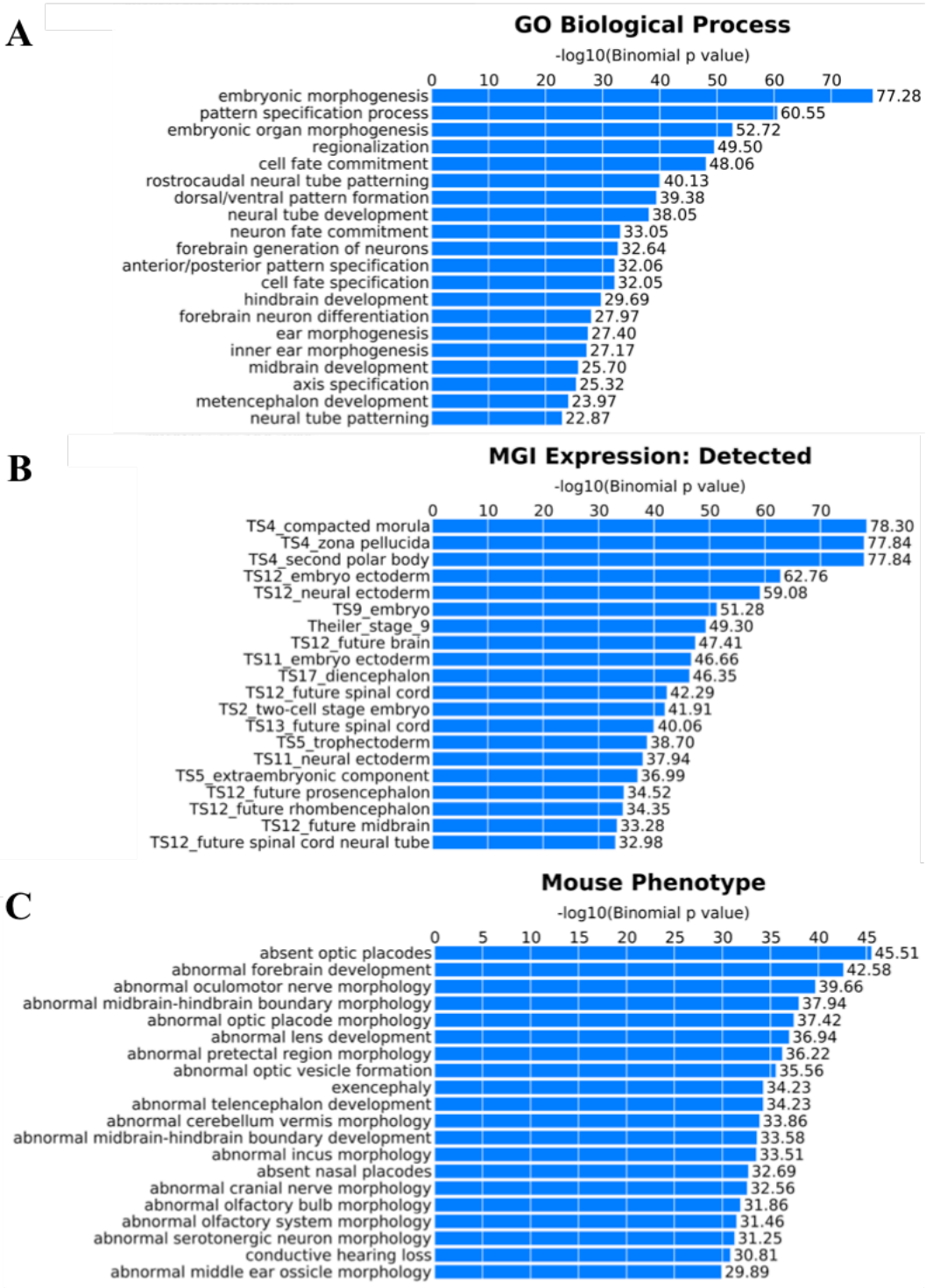


Figure 4-5: Highest-ranked GO biological process (A), tissue expression (B) and mouse phenotype (C) terms according to binomial p-value for the test set of hyper-DMR genomic regions using GREAT (v3.0.0). The coordinates of Hyper-DMRs (mm10 genome assembly) identified between mESCs and mNSCs were submitted to GREAT (v3.0.0) for functional annotation.

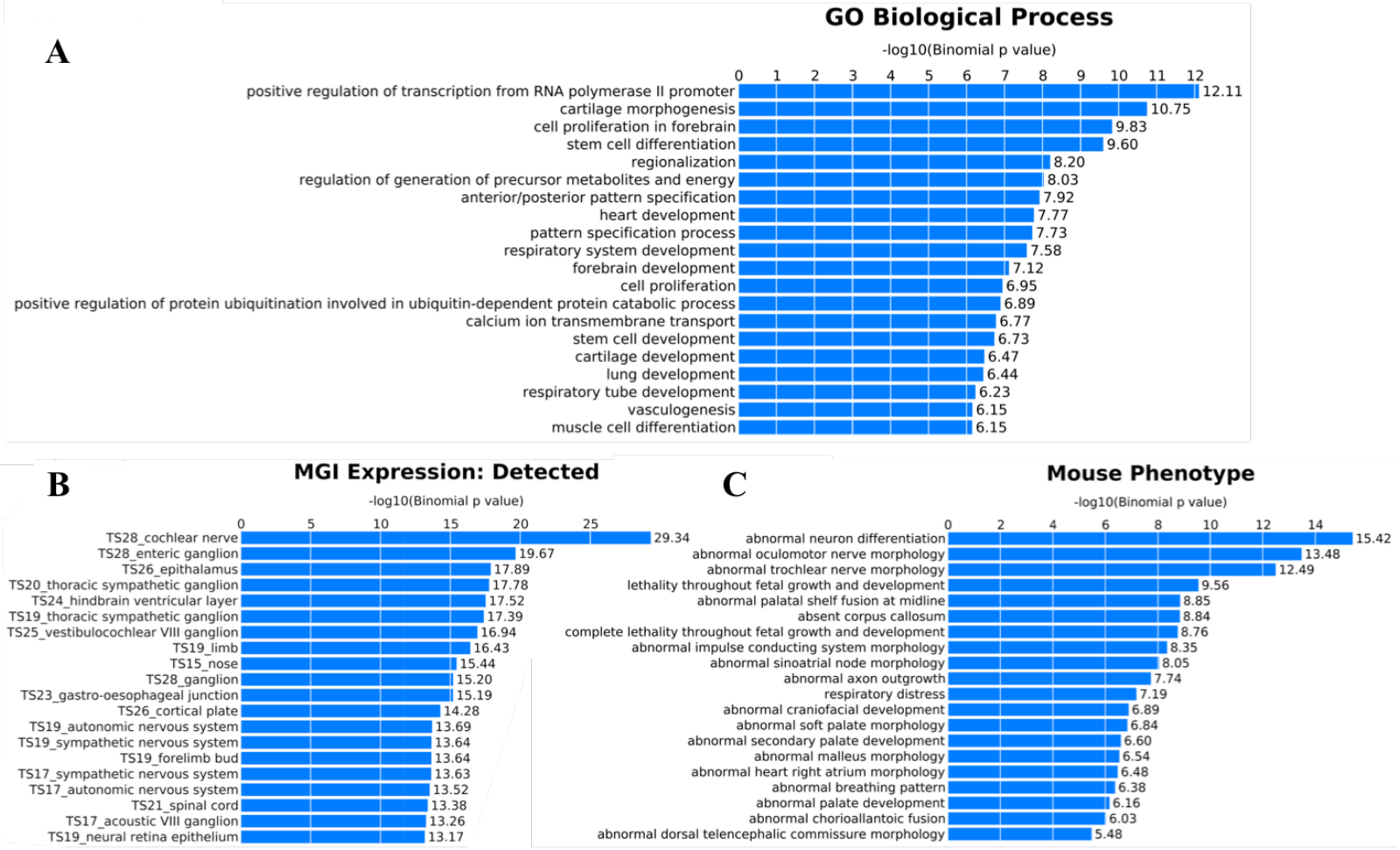


Figure 4-6: Highest-ranked GO biological process (A), tissue expression (B) and mouse phenotype (C) terms according to binomial p-value for the test set of hypo-DMR genomic regions using GREAT (v3.0.0). The coordinates of Hypo-DMRs (mm10 genome assembly) identified between mESCs and mNSCs were submitted to GREAT (v3.0.0) for functional annotation.

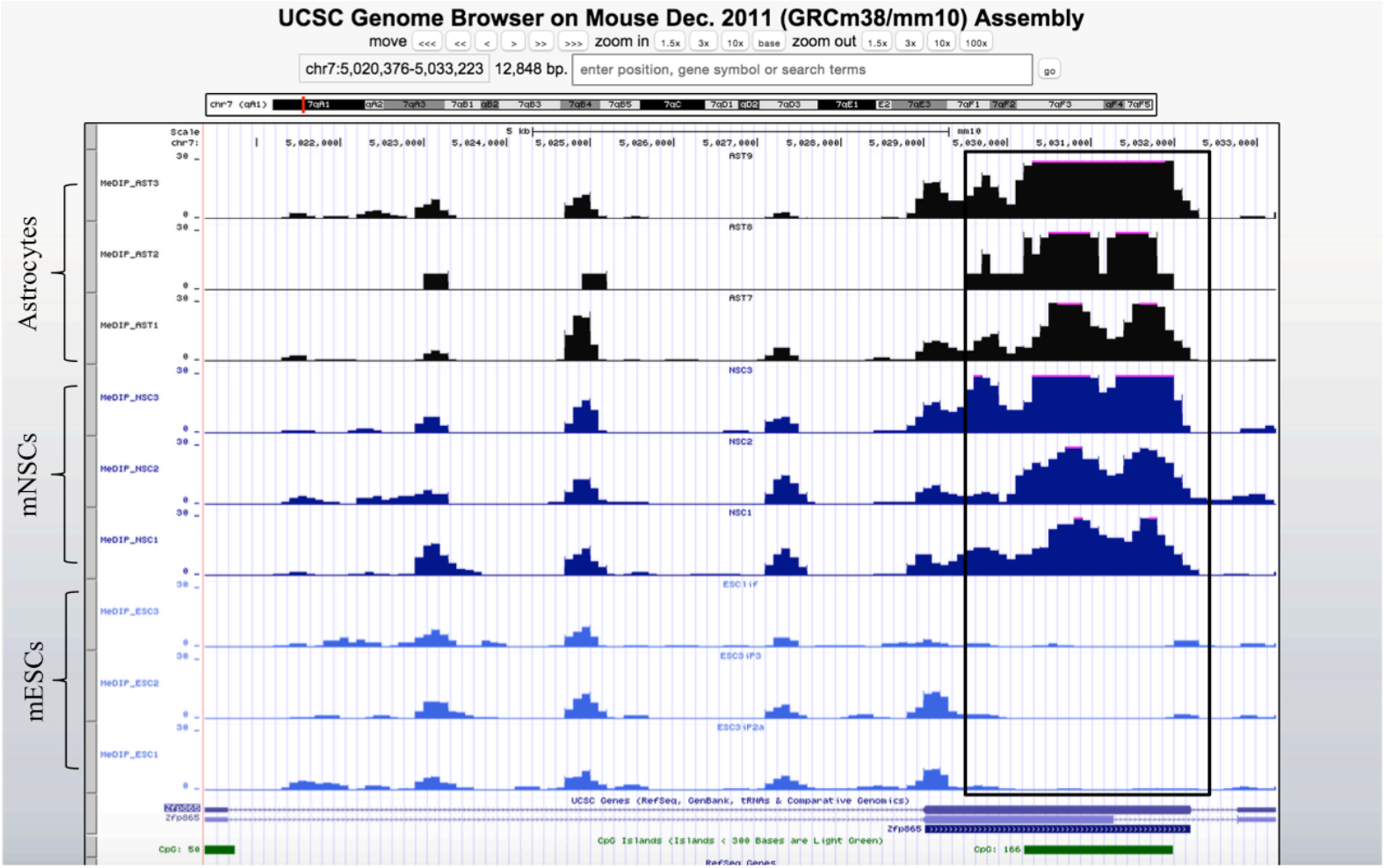


Figure 4-7: Example of a hypermethylated DMR called between mESCs and mNSCs. Wig tracks of all nine MeDIP-seq profiles were uploaded to the UCSC genome browser (mm10 build) to visualise DMRs. The black box (above) indicates the location of the most statistically significant hyper-DMR (Bonferroni-adjusted p-value = 7.01×10^{-29}), located within a CpG island in the final coding exon of the zinc finger protein 865 (*Zfp865*) gene located on chromosome 7. The region remains hypermethylated upon subsequent differentiation to astrocytes. Comparison to WGBS data available in Ensembl (Supplementary Figure 12) confirmed this region to be hypomethylated in mESCs and associated with both H3K4me3 and H3K27me3 modifications, indicating a poised chromatin state. In mNPCs, the region is fully methylated.

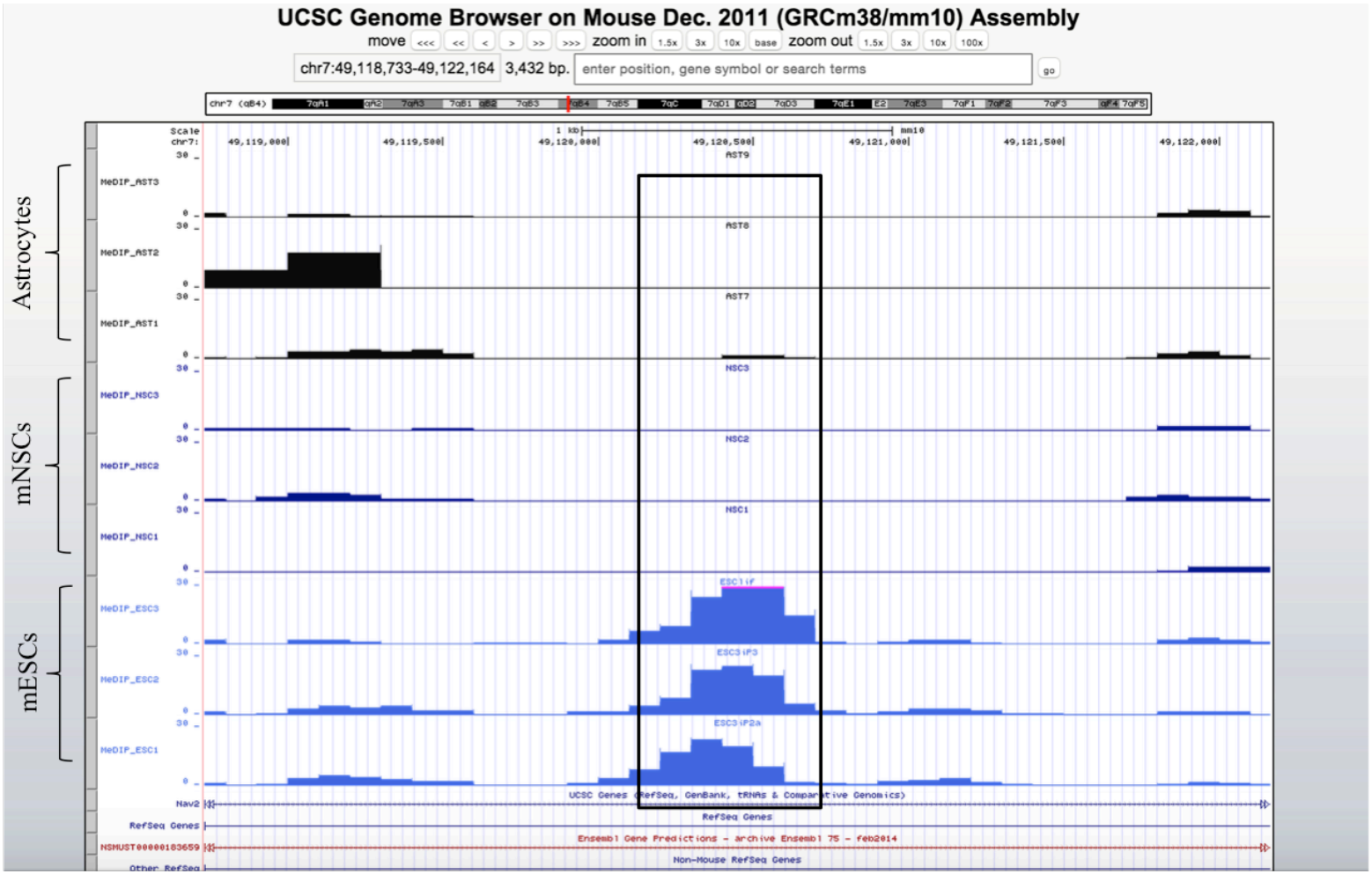


Figure 4-8: Example of a hypomethylated DMR called between mESCs and mNSCs. Wig tracks of all nine MeDIP-seq profiles were uploaded to the UCSC genome browser (mm10 build) to visualise DMRs. The black box (above) indicates the location of one of the most statistically significant hypo-DMRs (5th most significant overall, Bonferroni-adjusted p-value = 2.00 x 10^{-13}). This DMR sits within an intronic region of the neuron navigator 2 (*Nav2*) gene located on chromosome 7. Comparison to WGBS data available in Ensembl (Supplementary Figure 13) confirmed that this region is fully methylated in mESCs and associated with the repressive histone modification H3K27me3. The same region is substantially demethylated in mNPCs, supporting the MeDIP-seq data presented above.

4.5 Functional annotation of differentially hydroxymethylated regions

Functional annotation of the hypo- and hyper-hDMRs called between mESCs and mNSCs (Table 4-3) did not support a clear association between 5hmC changes and neural differentiation processes. Hyper-hDMRs submitted to GREAT revealed gene enrichment for four biological processes and five molecular function terms (Supplementary Figure 9), all related to cellular metabolism and respiratory function. Further analysis revealed that the majority hDMRs associated with these terms mapped to the mitochondrial genome (14 out of a total of 25 hyper-hDMRs) and are unlikely to be true regions of differential 5hmC content. Despite previous reports of mitochondrial DNA (mtDNA) methylation observed through MeDIP assays (Shock et al., 2011), base-resolution BS-seq analysis has ruled out the possibility of CpG methylation in mtDNA (Hong et al., 2013). The extremely high levels of both 5hmC and 5mC observed in mtDNA compared to autosomes through MeDIP- and hMeDIP-seq is perhaps due to an antibody immunoprecipitation bias towards mtDNA, which has a high copy number per cell relative to nuclear DNA chromosomes.

The most strongly enriched tissue expression term, palatal shelf, highlighted 20 hyper-hDMRs and 8 associated non-mitochondrial genes (including *En1*, *Satb2* and *Pax3*) with potential functional relevance during neural specification. *Satb2* encodes a protein involved in chromatin remodelling and transcription regulation through specific binding to nuclear matrix attachment regions and recruitment of histone-modifying enzymes (HDACs and HATs) to regulatory DNA sites (Britanova et al., 2005; Gyorgy et al., 2008). *Satb2* thought to play a critical role in early embryonic development through the regulation of *Nanog* expression and ESC pluripotency (Savarese et al., 2009). There is also evidence for a regulatory role in the initiation of upper-layer 1 neurons and downregulation of genes specific to deep-layer neurons in the developing neocortex (Britanova et al., 2008; Leone et al., 2014). Interestingly, whole genome bisulphite sequencing (WGBS) data of mESC and mNPC genomes (available as regulatory tracks on the Ensembl genome browser) shows that the *Satb2*-associated hyper-hDMR is partially methylated in mESCs but fully methylated in mNPCs (Figure 4-9). As 5mC and 5hmC are indistinguishable after BS conversion, the hMeDIP-seq data suggests that this apparent increase in 5mC

between mESCs and mNPCs may in fact be due to gains in 5hmC rather than 5mC and may represent a region undergoing demethylation in differentiating mESCs.

Further analysis of the hyper-hDMR associated gene set using the DAVID online tool found enrichment for 40 GO terms including cell projection, neurogenesis and multicellular organism development (Supplementary Figure 7). The most highly enriched tissue expression term corresponded to genes upregulated in brain (p-value = 4.5 $x10^{-2}$; Supplementary Figure 8). One of these genes, *Mid1*, encodes a protein involved in microtubule dynamics and body axis patterning. The associated hyper-hDMR was the most statistically significant region identified between mESC and mNSC samples which did not fall within repeat DNA sequence (Figure 4-10) and is located approximately 2kb downstream of *Mid1*. Further analysis using the regulatory feature tracks available for the mm10 genome assembly in Ensembl revealed that this hDMR sits over a DNase1 hypersensitive site and is marked by both active and repressive histone modifications (H3K4me3 and H3K27me3, respectively) in mESCs, but not mNPCs, suggesting that this may be an important regulatory region (Supplementary Figure 14). It appears that this region exists in a poised chromatin state in mESCs in the absence of either 5mC or 5hmC and appears to gain 5hmC during neural differentiation, concomitant with a loss of all histone modifications. However, WGBS data available on the Ensembl browser suggests that this region is partially methylated in both mESCs and mNPCs, contradicting the MeDIP-seq data presented here, which found the region to be unmethylated in both mESCs and mNSCs. It is possible that the either 5mC signal was underestimated in the MeDIP libraries due to limited sequencing depth or that the WGBS data overestimated the proportion of 5mC within this region due to the inability to distinguish 5mC from 5hmC after BS conversion.

In contrast, GREAT analysis of hypo-hDMRs found no significant enrichment for particular GO annotations. However, analysis of this associated gene set (defined as genes within 100kb upstream or 50kb downstream of a particular hDMR) using the DAVID online tool found enrichment for a handful of GO terms, including for genes upregulated in head tissue and genes involved in cell communication and signal transduction (Supplementary Figure 10). Further investigation into the hDMR

regions associated with genes upregulated in head tissue (*Vegfc, Dopey2, Aida, Abcb10* and *Tmem196*) using the UCSC genome browser revealed that all five regions mapped to repetitive DNA sequence and were unlikely to represent true 5hmC signal. Moreover, just two of the 119 identified hypo-hDMRs did not overlap any repeat elements, highlighting the difficulty in assessing whether or not real differences in 5hmC distribution were detected through hMeDIP-seq profiling. The strong bias in sequence coverage towards repetitive DNA is likely due to a combination of partial cross-reactivity of the anti-5hmC antibody to cytosine-rich DNA sequence (e.g. CA/CT simple repeats) and the extremely low abundance of 5hmC in genomic DNA, perhaps making immunoprecipitation of hydroxymethylated fragments much less efficient than seen for 5mC in MeDIP reactions.

Nevertheless, the most statistically significant hypo-hDMR (Bonferroni-adjusted p-value = 2.59×10^{-21}) detected through the MEDIPS analysis pipeline corresponded to one of the two hypo-hDMRs falling within non-repetitive sequence (Figure 4-11), located 1kb upstream of *Atpase6*, a mitrochondrial protein-coding gene. This hDMR overlaps regulatory DNA sequence characterised by DNase1 hypersensitivity in mESCs and Pol-II enrichment in mNPCs (Supplementary Figure 15), indicating that 5hmC dynamics at this region may have functional importance.

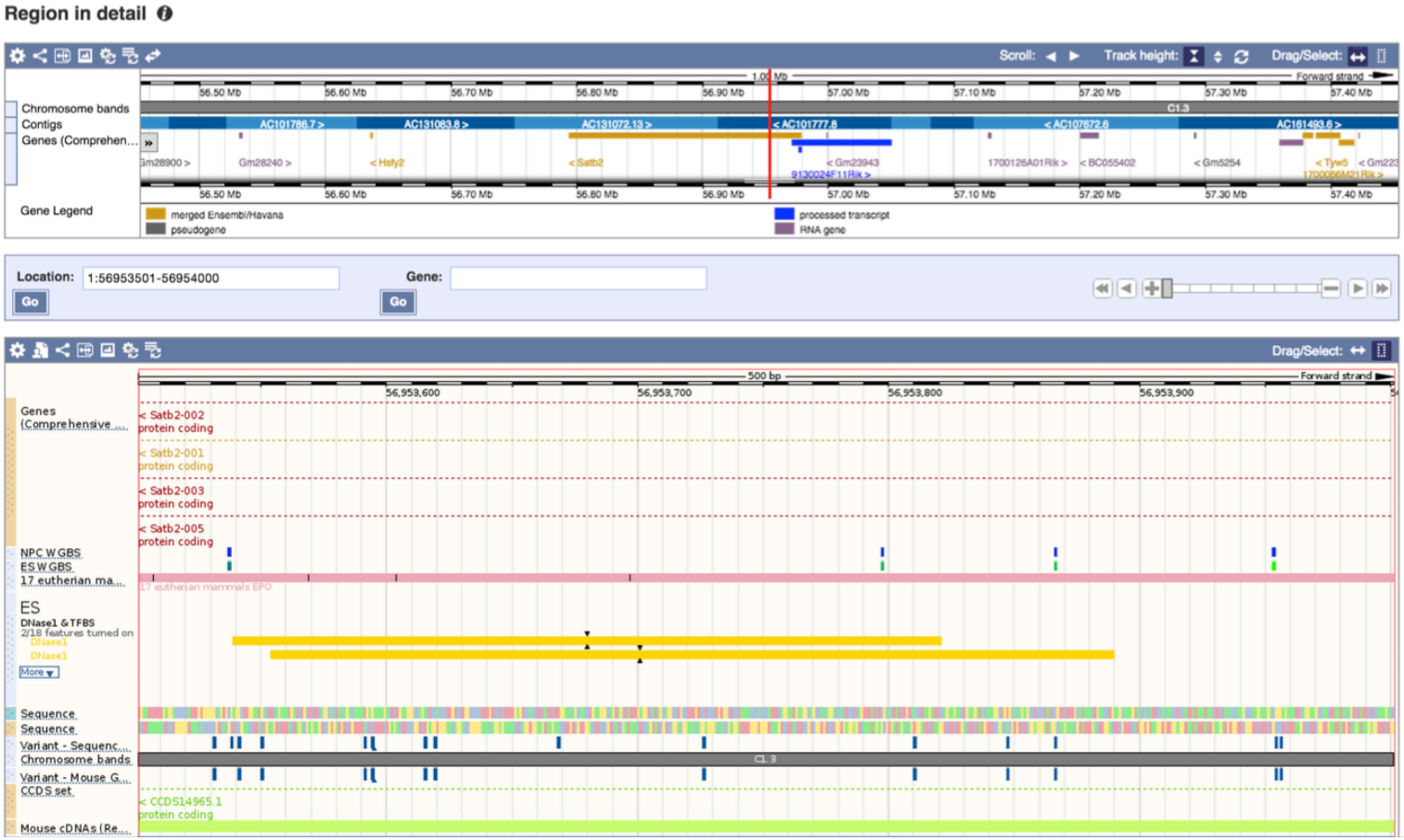


Figure 4-9: Ensembl genome browser of chr1:56953501-56954000, a 500bp window within an intronic region of the *Satb2* gene. This region was identified as a hyper-hDMR between mESCs and mNSCs (unmethylated in mESCs and hydroxymethylated in mNSCs) through hMeDIP-seq. WGBS tracks of mESCs and mNPCs (labelled as "ES WGBS" and "NPC WGBS" above) show this region is partially methylated in mESCs (45-75% 5mC/5hmC) and fully methylated in mNPCs

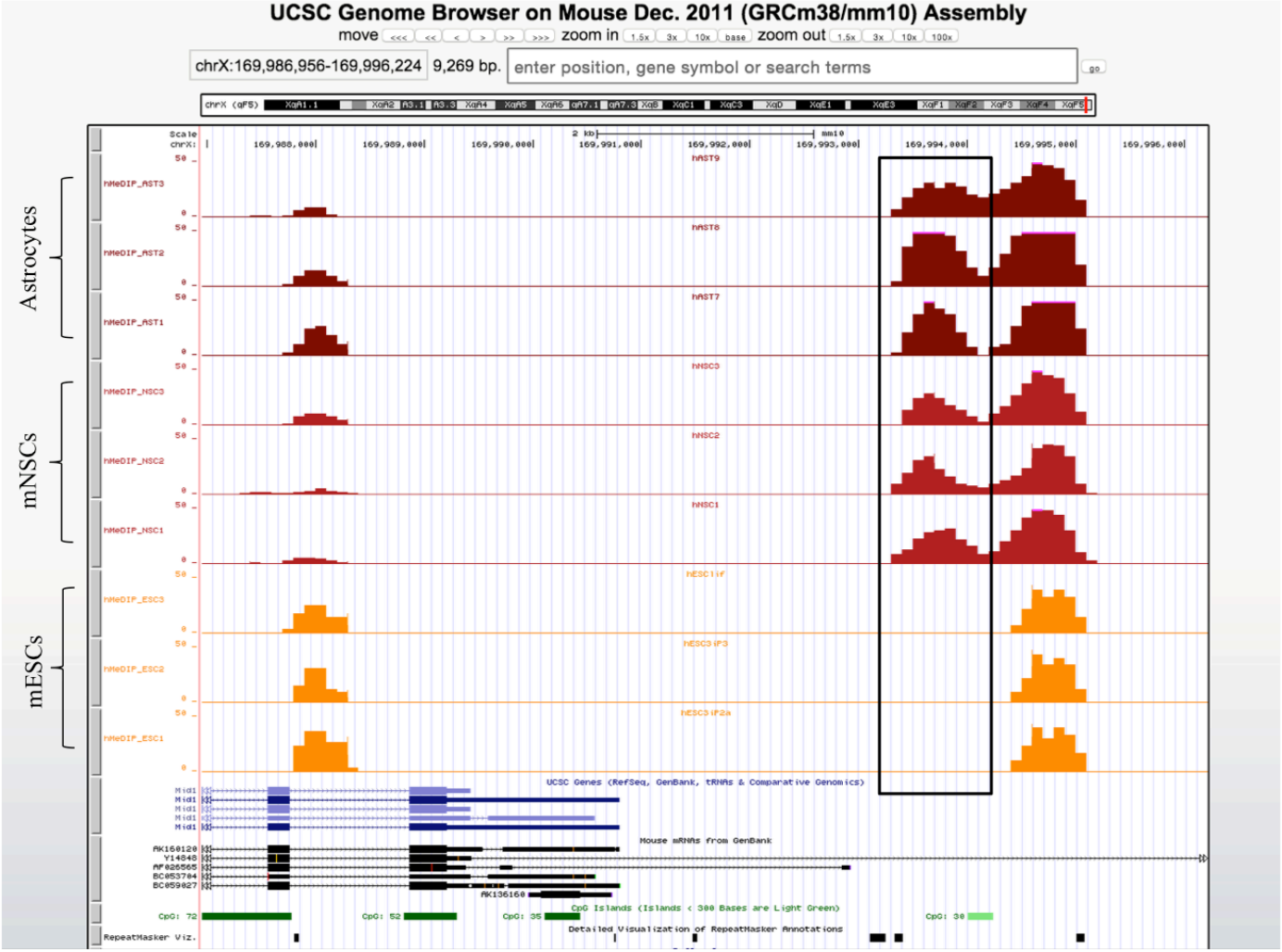


Figure 4-10: Example of a hyper-hDMR called between mESCs and mNSCs. Wig tracks of all nine hMeDIP-seq profiles were uploaded to the UCSC genome browser (mm10 build) to visualise hDMRs. The black box (above) indicates the location of the most statistically significant hyper-hDMR that did not fall within either repeat elements or the mitochondrial chromosome (6^{th} most significant overall, Bonferroni-adjusted p-value = 1.49×10^{-33}), located 2kb downstream of *Mid1*.

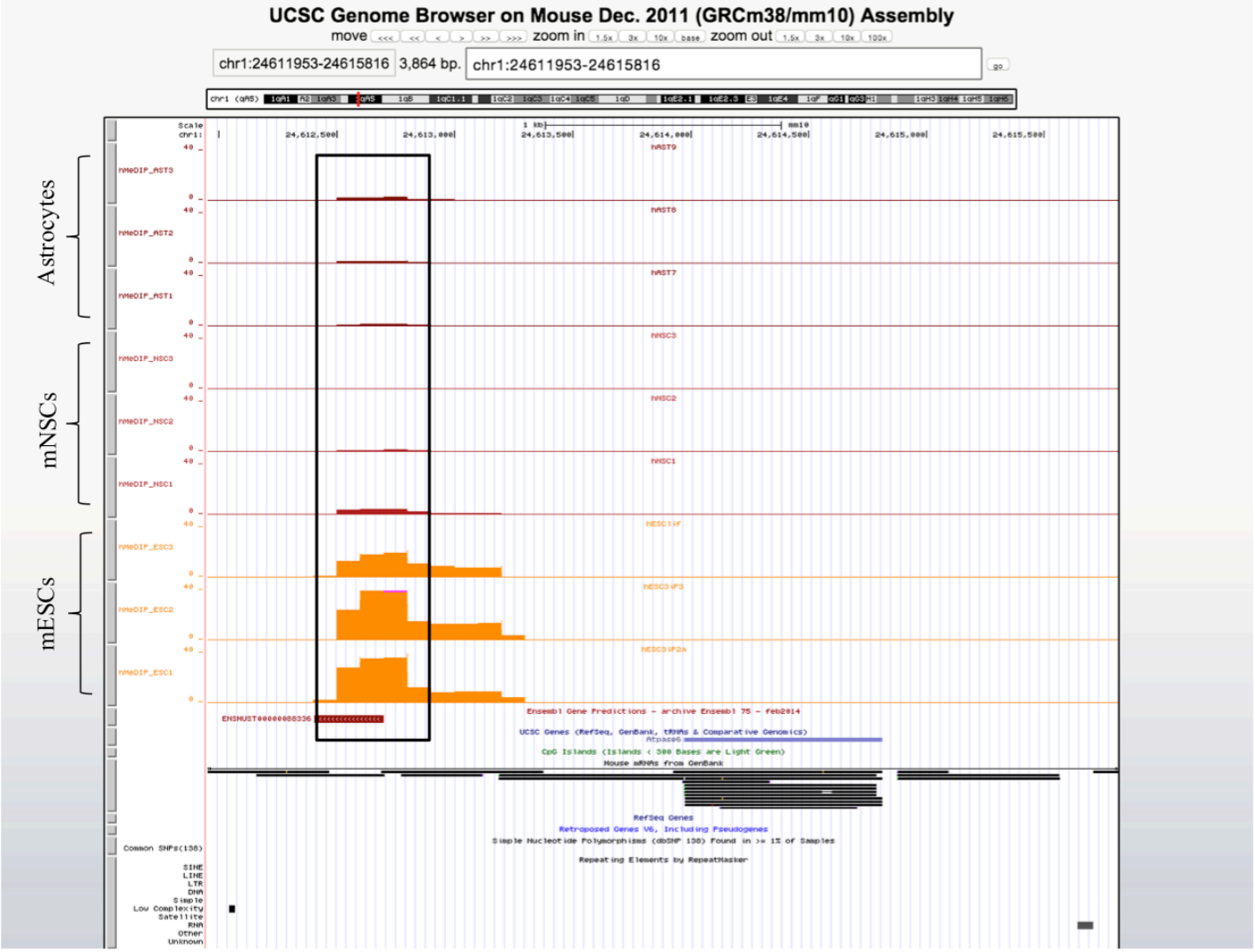


Figure 4-11: Example of a hypo-hDMR identified during the mESC to mNSC transition. Wig tracks of all nine hMeDIP-seq profiles were uploaded to the UCSC genome browser (mm10 build) to visualise hDMRs. The black box (above) indicates the location of the most statistically significant hypo-hDMR that did not fall over repeat elements (2^{nd} most significant overall, Bonferroni-adjusted p-value = 2.59×10^{-21}), located approximately 1kb upstream of Atpase6.

4.6 Validation of MeDIP- and hMeDIP-seq profiles using targeted oxBS-pyrosequencing

Due to concerns over the high proportion of repeat DNA sequence captured through the hMeDIP-seq analysis and inconsistencies with other publically available genome-wide 5hmC profiles, oxidative bisulphite conversion was adapted for use with pyrosequencing technology (oxBS-pyroseq) in order to validate the observed changes in 5mC and 5hmC during neural differentiation. This method allows single-base resolution of both 5mC and 5hmC levels at individual CpG sites through the comparison of BS- and oxBS-converted DNA pyrosequencing results and was performed in a locus-specific manner to assess the absolute abundance of 5mC and 5hmC at particular DMR and hDMRs of interest.

First, in order to rule out the possibility that the hMeDIP reaction resulted in unspecific pull-down of genomic DNA fragments rather than specific enrichment of 5hmC-containing DNA, four randomised hMeDIP datasets were generated and compared to all 18 MeDIP- and hMeDIP-seq profiles using unsupervised hierarchical clustering of their methylation score. Heatmap representation of the resulting clusters shows a distinct, inverse enrichment pattern for the hMeDIP datasets compared to MeDIP (Figure 4-12A). The four randomly-generated profiles clustered separately from all other samples and show uniform enrichment levels across all genomic features (CpG islands, CpG shores, exons, introns, intergenic regions, promoters and gene transcripts), although their profiles do show greater similarity to the enrichment patterns observed in hMeDIP profiles than those of MeDIP (Figure 4-12B). Taken together, this analysis suggests that all sequenced DNA libraries underwent a specific and non-random selection step during DNA immunoprecipitation, although the background signal in the hMeDIP profiles appears to be very high compared to the MeDIP data.

Next, pyrosequencing assays were designed for eight individual hDMR regions (all identified between mESCs and mNSCs) and tested in triplicate on both oxBS- and BS-converted DNA from four biological samples: *Sox1-GFP* E14 mESCs, mNPCs (Sox1+), mNSCs (Sox1-) and differentiated astrocytes. The *in vitro* derivation of these samples is described fully in chapter 3, section 3.4.

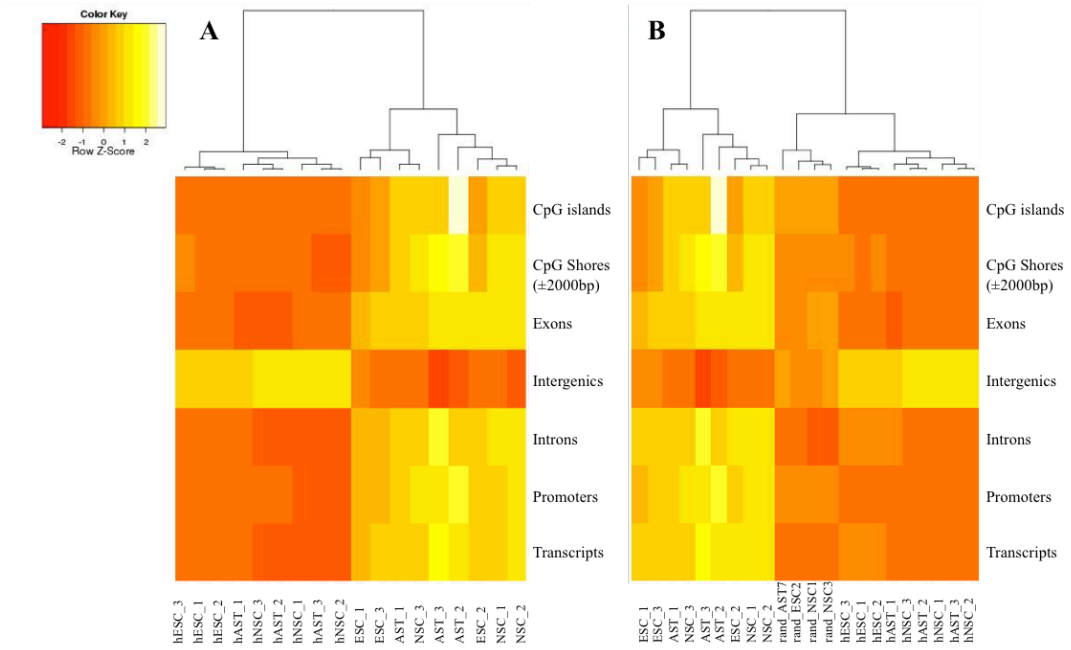


Figure 4-12: Heatmap representation of unsupervised hierarchical clustering of all MeDIP and hMeDIP profiles according to their methylation enrichment scores (Z-scores) separate into two distinct groups. A) hMeDIP profiles cluster as a distinct group (left) and show low enrichment in intragenic genomic features such as introns, exons and promoters. In contrast, MeDIP profiles (right) show low enrichment in intergenic regions but are enriched for 5mC across all other features. B) Inclusion of 4 randomly-generated genome-wide methylation profiles into the analysis shows that both MeDIP and hMeDIP profiles remain clustered as two independent groups distinct from these controls (centre), which showed relatively uniform enrichment across all genomic features. This suggests that the DNA immunoprecipitation step during MeDIP and hMeDIP library preparation successfully resulted in the specific and non-random selection of fragmented DNA.

The levels of 5hmC at each region were determined through subtraction of the oxBS-pyroseq assay methylation scores from the BS-pyroseq assay methylation scores. However, no substantial differences in 5hmC were recorded between the mESC and mNSC samples using this assay at any of the eight regions. In addition, 5hmC levels were uniformly low across regions in all samples, ranging from 0 – 15.7% 5hmC. For example, validation of a hyper-hDMR called between mNSCs and mESCs revealed only a 4% increase in 5hmC levels through oxBS-pyrosequencing, from 2.7 to 6.7% 5hmC (Figure 4-13A). In contrast, an oxBS-pyroseq assay designed around a hyper-DMR region identified between mNSCs and mESCs unambiguously confirmed the observed gain in 5mC (Figure 4-13B). The region, which appeared fully unmethylated in mESCs and substantially methylated in mNSCS according to MeDIP-seq profiling (Supplementary Figure 16), was found to contain just 5% 5mC in mESCs by oxBS-pyroseq, subsequently rising to 50% 5mC in mNSCs (Sox1-).

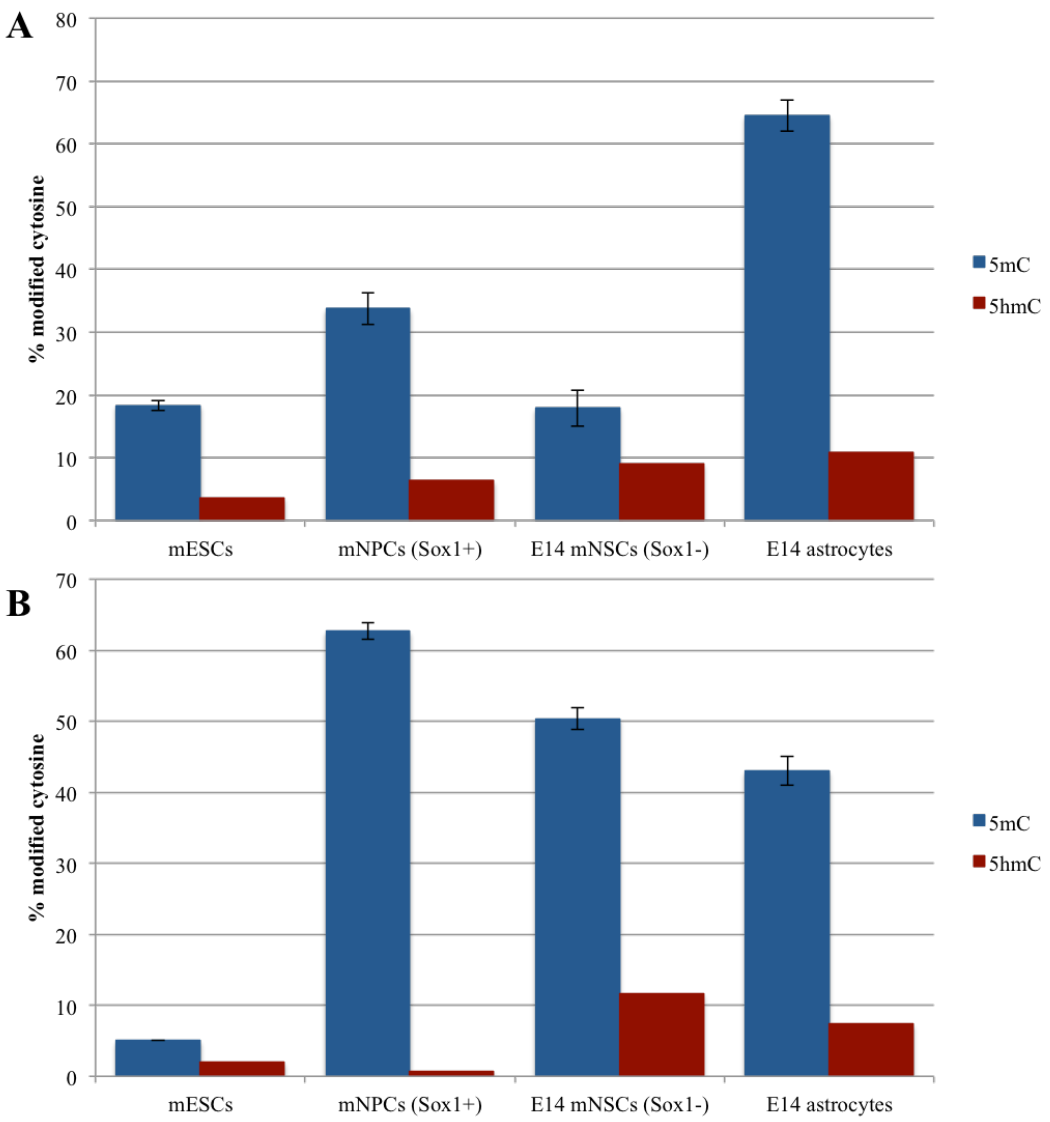


Figure 4-13: Pyrosequencing validation of 5mC and 5hmC levels at two genomic regions in *Sox1-GFP* mESCs, mNSCs and astrocytes. Sox1+ mNPCs represent an intermediary, rosette-forming stage during the differentiation of mESCs into stable mNSC lines. A) Validation of a hyper-hDMR identified between mNSCs and mESCs located at chr17:27,422,001- 27,422,500, at the 3' end of Grm4, using an oxBS-pyroseq assay. The assay interrogated a 225bp region located at chr17:27,422,334-27,422,558. Despite substantial differences in 5hmC levels identified by hMeDIP-seq across this region, only a small increase in 5hmC from 3.6 to 9.0% was observed between mESCs and mNSCs (Sox1-). 5mC levels in this region remained relatively constant during this transition, at approximately 18% 5mC, but increased to 64.5% 5mC upon differentiation into astrocytes. Error bars represent the standard deviation between triplicate oxBS-pyroseq assay results. B) Validation of a hyper-DMR identified between mNSCs and mESCs located at chr13:44,705,001-44,706,000 using an oxBS-pyroseq assay. This hyper-DMR falls in an intergenic region approximately 20kb upstream of Jarid2. The oxBS-pyroseq assay interrogated a 179bp sequence located at chr13:44,705,585-44,705,763 and results confirmed that this region is hypermethylated in mNSCs compared to mESCs, with a methylation difference of 45% measured between the two sample types. This region remains substantially methylated upon subsequent differentiation into astrocytes, although a small decrease of 7% 5mC is seen upon mNSC differentiation.

4.7 Comparison of global DNA 5mC and 5hmC levels between cultured and primary cells

Genome-wide profiles based on immunoprecipitation techniques, such as MeDIP- and ChIP-seq, provide data only on the relative enrichment and distribution of a particular modification and cannot be used to calculate absolute levels within each sample. In order to assess whether 5hmC levels had increased significantly during neural differentiation, as had been previously predicted, global 5hmC levels were quantified using a sensitive liquid chromatography-mass spectrometry (LC-MS) method across the three defined cellular differentiation time points (mESCs, mNSCs and astrocytes) and compared to adult mouse brain tissue.

Duplicate measurements of 5hmC and 5mC levels were made across 10 cultured cell samples, including three mESC samples, four mNSC samples, two *in-vitro* differentiated astrocytes samples and one *in-vitro* differentiated neuron-like sample (see Chapter 3 for details of *in vitro* cell samples). Surprisingly, very little change in global 5hmC levels was observed between mESCs, mNSCs and differentiated neural cells (Figure 4-14). Although a large increase in 5hmC content had been expected across these three stages of neural differentiation, both wild-type C57BL/6N and *Sox1-GFP* E14 mESCs had similar levels of 5hmC as differentiated astrocytes and neurons, in the range of 0.06 - 0.08% of total cytosine. A decrease in 5hmC levels of over 50% was noted in all mNSCs in comparison to mESCs, regardless of whether they were primary mNSCs ("WT C57 mNSCs") or *in-vitro* differentiated mNSCs ("*Sox1-GFP* mNSCs"), falling within the range of 0.02 – 0.04%.

In contrast, 5hmC measurements in DNA extracted from various adult mouse brain tissue samples revealed levels substantially higher (approximately 10-fold) than those in all *in vitro* cultured cell samples. 5hmC levels in four separate cerebellum samples fell within the range of 0.36 – 0.40% whilst 5hmC in the cortex, olfactory bulb and other unclassified brain tissue was found at even higher levels, in the range of 0.58 - 0.67%.

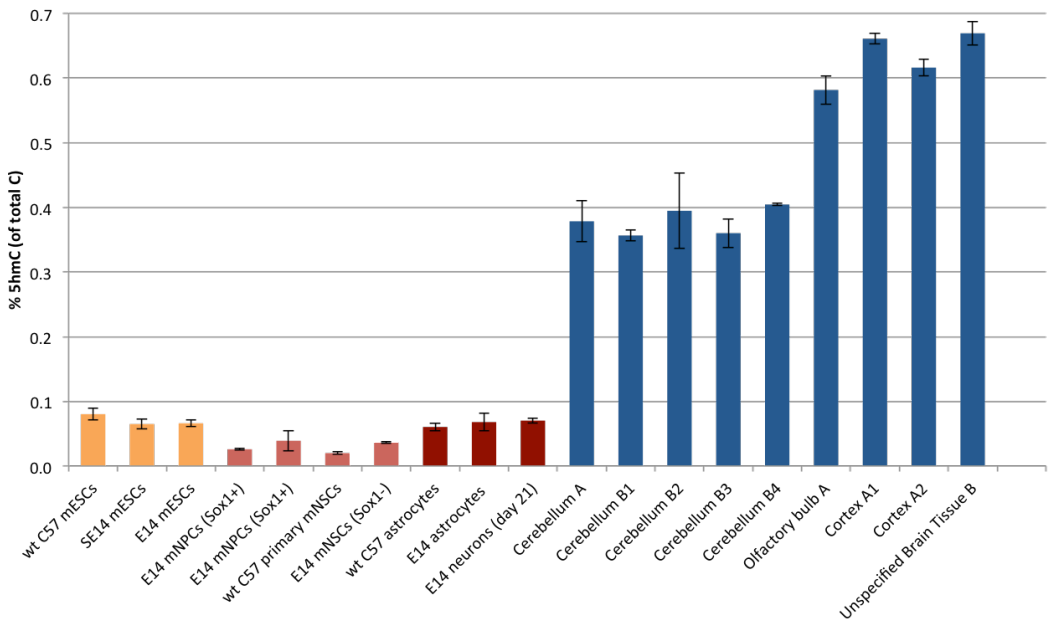


Figure 4-14: LC-MS quantification of the global 5hmC content of cultured cells and adult tissues. 5hmC levels are plotted as a percentage of the total cytosine content of each DNA sample; each bar represents the mean value of duplicate LC-MS measurements. Error bars represent the standard deviation. WT C57 cells refer to those samples used for MeDIP/hMeDIP-seq profiling. All E14 cells were derived *in vitro* from a *Sox1-GFP* E14 cell line (see Chapter 3, section 3.4). Mouse tissue samples were dissected from two 6-month old female mice (A and B, respectively).

Despite striking differences in 5hmC content between cultured cell lines and primary tissue samples, the same pattern was not observed for 5mC (Figure 4-15). Global 5mC levels showed some variation across samples, covering the range of 2.82 – 4.23% of total cytosine, but did not show significant differences between cultured cells and primary tissues. There was also a fair amount of variation within biological replicates; two *Sox1-GFP* E14 astrocyte samples showed a difference of 0.73% in their total 5mC content, whilst two *Sox1-GFP* E14 mNPC samples (Sox1+ cells collected at day 10 of mESC differentiation; see Section 3.4 for derivation) showed a difference of 0.38%.

There did not appear to be a direct inverse correlation between 5hmC and 5mC levels, as would be expected if 5hmC was a stable by-product of DNA demethylation; the small increase in 5hmC (~0.03%) between mNSCs and differentiated astrocytes, for example, was not accompanied by an equivalent reduction in 5mC levels. In fact, average 5mC levels decreased by 0.38% during this transition. If 5hmC is truly an intermediary product in a DNA demethylation pathway, it appears likely that the majority of 5hmC, as least within proliferating

cells in culture, rapidly undergoes further conversion to 5fC, 5caC or fully unmodified cytosine (not measured), making it difficult to capture the true modified cytosine content at a particular developmental stage. In addition, balanced gains and losses in either 5hmC or 5mC occurring throughout the genome would not be reflected in these global measurements, masking the existence of dynamic but equivalent changes in their distribution.

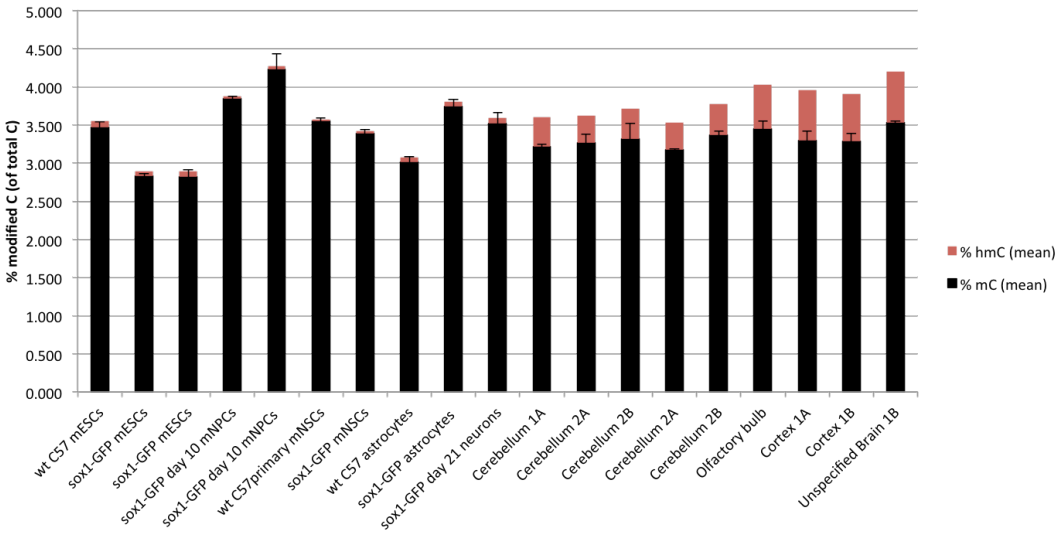


Figure 4-15: LC-MS quantification of the combined 5mC and 5hmC content of various mouse DNA samples. Black bars indicate the total 5mC content and red bars indicate total 5hmC content of each sample type as a percentage of total cytosine. Each bar represents the mean value of duplicate LC-MS measurements. Error bars represent the standard deviation of the 5mC measurements only. WT C57 cells refer to those samples used for MeDIP/hMeDIP-seq profiling. All E14 cells were derived *in vitro* from a *Sox1-GFP* E14 cell line (see Chapter 3, section 3.4). Mouse tissue samples were dissected from two 6-month old female mice (A and B, respectively).

4.8 Discussion

Genome-wide changes in 5mC and 5hmC were mapped between three defined cellular stages of a neural differentiation pathway in order to shed light on the dynamic relationship between these two cytosine modifications and the potential functional relevance of 5hmC. A general trend of increasing locus-specific gains in both 5mC and 5hmC was observed during the differentiation process, with most of these changes occurring early on in development, between the embryonic and neural stem cell stage. However, this pattern was not reflected by global changes in either 5mC or 5hmC content. 5mC levels remained relatively constant across all samples, with both wild-type C57 mESCs and primary mNSCs containing 3.5% 5mC (of total cytosine), whilst differentiated astrocytes contained 3.4% 5mC. Conversely, 5hmC content decreased by approximately 50% between mESCs and mNSCs, despite over 60% of the identified hDMRs during this transition corresponding to gains in 5hmC.

These discrepancies highlight the fact that most immunoprecipitation-based genome profiling methods are not quantitative and cannot be used to infer absolute differences of their respective modifications. Interestingly, a recent study describes a novel comparative hMeDIP-seq method that employs barcode technology to perform antibody-based immunoprecipitation on multiple DNA samples in one reaction, thereby preserving quantitative differences in 5hmC between samples (Tan et al., 2013). Using this method, they observed a 78% decrease in global 5hmC between mESCs and *in vitro* differentiated mNPCs, consistent with the LC-MS measurement data presented here, suggesting that embryonic and neural stem cells are characterised by distinct hydroxymethylation states that likely hold biological significance.

Comparison between DMR and hDMR regions did not reveal any overlap in the location of observed 5mC and 5hmC changes. Although pre-existing 5mC is required for 5hmC generation via TET-mediated oxidation, sites of hyper-hydroxymethylation did not coincide with any hypo-DMRs, as would be expected. However, this is probably due to underestimation of the total number of DMR and hDMR sites within the genome as a result of low sequence coverage. In addition, the kinetics of 5mC oxidation may simply occur too quickly to efficiently capture

intermediary products (5hmC, 5fC or 5caC) at sites of active demethylation, making it difficult to pinpoint regions undergoing dynamic hydroxymethylation changes. It appears that the hMeDIP assay was not sensitive enough on this occasion to detect more subtle differences in 5hmC over background signal. Further optimisation of the IP reaction in combination with much deeper sequencing is probably needed in order to draw definitive conclusions concerning 5hmC dynamics between samples.

Analysis of the genomic distribution of hDMRs revealed that over 80% fell within repetitive DNA elements, more than would be expected by chance alone. This bias towards repeat sequence may reflect cross-reactivity of the anti-5hmC antibody to high-density regions of unmethylated cytosine. The extremely low abundance of 5hmC in the genomic DNA samples being profiled combined with the tendency for 5hmC to be located in regions of moderate-to-low CpG density (Yu et al., 2012) implies that immunoprecipitation efficiency is also likely to be much lower than that observed for 5mC in traditional MeDIP-seq reactions.

Only a small percentage of the identified hDMRs between mESC, mSNCs and astrocytes seemed to represent true differences in 5hmC distribution, due to the apparent antibody bias towards cytosine-rich repeat regions of the genome. Attempted validation of hMeDIP-seq profiles using an alternative sequencing-based technology, oxBS-pyroseq, failed to confirm the presence of differential 5hmC levels at selected hDMR sites. However, this assay was constrained by the fact that pyrosequencing primer annealing sites could not fall within repeat DNA sequence or CpG loci, severely restricting the number of hDMRs that could be successfully interrogated. In addition, each oxBS-pryoseq assay only measured the methylation status of one or two CpG sites at a time due to the low CpG density of hDMRs. As a consequence, individual pyrosequencing assays are potentially unrepresentative of the 5hmC content of the corresponding hDMR as a whole.

Despite fluctuations in global 5hmC levels between mESCs, mNSCs and astrocytes, 5hmC content nevertheless remained very low (0.02-0.08% of total cytosine) across all cultured cell samples and did not increase substantially in differentiated astrocytes, as had initially been expected. 5hmC levels in cultured cell lines were at least 4-fold lower than 5hmC levels measured in adult mouse brain tissues, which

fell between 0.36 – 0.67% of total cytosine. It appears unlikely that low 5hmC levels are a constitutive element of astrocytes; several other studies have since described similarly low 5hmC levels across a wide range of cell lines. Decreased 5hmC appears to be a general feature of cell adaptation to culture conditions and is not restricted to particular cell types (Nestor et al., 2012).

A 40% reduction in global 5hmC content in *in vitro* differentiated dopamine neurons compared to human ESCs was recently reported (Kim et al., 2014), contradicting earlier findings of exceptionally high 5hmC levels in primary neurons isolated from an adult mouse brain (Kriaucionis and Heintz, 2009). In addition, immunohistochemical analysis of foetal, paediatric and adult human brain sections demonstrated that all terminally differentiated neural cell types (neurons, astrocytes and oligodendroglial cells) show high levels of 5hmC staining (Orr et al., 2012), whilst progenitor cell populations show relatively low 5hmC staining. It therefore appears that *in vitro* differentiation protocols for the generation of neural subtypes do not recapitulate 5hmC states observed *in vivo* and suggests that further analysis of 5mC/5hmC dynamics during neurodevelopment may be best addressed using primary cell and tissue samples.

Chapter 5 Novel methods for DNA methylation studies

Work carried out in collaboration:

The human frontal cortex DNA sample was kindly provided by Dr Harry Bulstrode (UCL Cancer Institute, University College London). Martin Bachman (Department of Chemistry, University of Cambridge) carried out all LC-MS analyses of genomic DNA samples. Mikiembo Kukwikila and Ha Phuong Nguyen (Department of Chemistry, University College London) carried out the synthesis, purification and HPLC analysis of the two caged nucleoside analogues, under the supervision of Dr Stefan Howorka.

5.1 Introduction

In this chapter, two new methods are presented for use in studies of DNA methylation. The first concerns a novel approach for the genome-wide profiling of 5hmC alongside 5mC using a well-established DNA methylation microarray, the Infinium HumanMethylation450 BeadChip. The second concerns a proof of principle study for the use of light-sensitive nucleoside analogues to achieve targeted demethylation events in a living system, with the potential for use in functional methylation studies and ultimately for therapeutic purposes.

5.2 An array-based method for genome-wide 5hmC analysis

5.2.1 Introduction

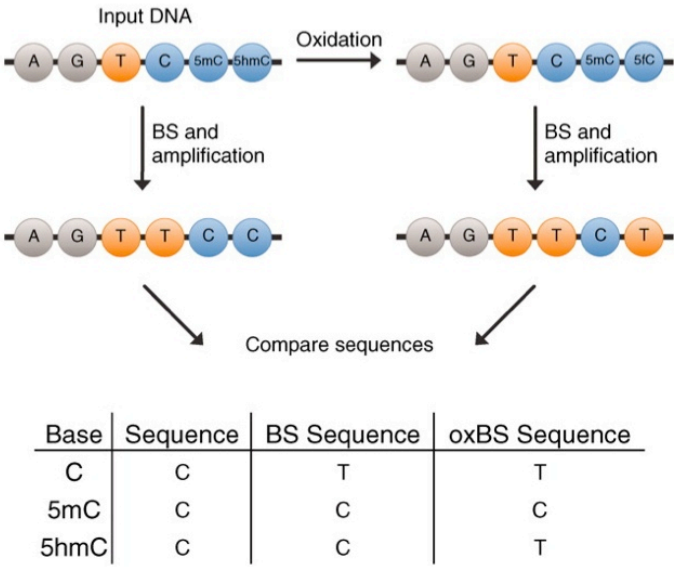
Current methods for profiling 5hmC at a genome-wide level rely largely on sequencing-based protocols that are both costly and time-consuming. The recent developments of both TET-assisted and oxidative bisulphite (oxBS) technologies have allowed single-base-resolution sequencing of 5hmC (Booth et al., 2012; Yu et al., 2012a), although the costs associated with whole-genome sequencing prevent widespread use of this technology, especially for projects depending on large sample cohorts. Immunoprecipitation-based methods for capturing hydroxymethylated DNA fragments represent an alternative approach to profiling DNA when coupled to sequencing, which substantially reduce the costs involved. However, these methods sacrifice single-base-resolution of 5hmC and are semi-quantitative at best. In

addition, antibody-based methods such as hMeDIP introduce bias in the 5hmC profile towards particular genomic regions due to preferential binding of the antibody to 5hmC in certain sequence contexts, including CpG-dense regions and repeat sequences (Matarese et al., 2011; Song et al., 2012).

Sequencing-based methods all require a high degree of expertise both for successful implementation of the chosen assay and downstream data analysis. Hence, there is a need for novel, readily accessible methods to analyse 5hmC on a genome-wide scale as research interest into the role of 5hmC expands. The Infinium HumanMethylation450 BeadChip is a widely-used, robust and reliable tool for DNA methylation profiling (Bibikova et al., 2011; Sandoval et al., 2011), suitable for high-throughput sample processing and is therefore the platform of choice for current epigenome-wide association studies (Michels et al., 2013; Rakyan et al., 2011). The 450K array interrogates over 480,000 CpG sites across the human genome, making it an appropriate tool for 5hmC detection as this modification has been reported to occur exclusively in the CpG context in both mammalian embryonic stem cells and frontal cortex tissue (Booth et al., 2012; Lister et al., 2013; Yu et al., 2012b). However, current workflows for this microarray rely on sodium bisulphite conversion of DNA, which discriminates between methylated and unmethylated cytosine bases within the genome but cannot differentiate between 5mC and 5hmC (Huang et al., 2010; Figure 5-1).

The availability of oxBS chemistry presents an opportunity to adapt BS-based 5mC profiling technology for 5hmC detection (Booth et al., 2012). Treatment of DNA with potassium perruthenate ($KRuO_4$) selectively oxidises 5hmC to 5fC, whereas both 5mC and C show no reactivity with the oxidant and remain unchanged. Traditional BS conversion follows the oxidation step; 5fC behaves in a manner analogous to unmodified cytosine under BS conditions, allowing only true 5mC positions to be detected as cytosine after oxBS conversion and PCR amplification of the DNA sample. Subtraction of oxBS-generated methylation profiles from standard BS-only methylation profiles allows for the detection of hydroxymethylated cytosine positions within the genome. A more accurate 5mC profile can be obtained by removing the confounding factor of 5hmC, whilst traditional BS conversion leads to

potential overestimation of 5mC levels in a given sample due to the presence of 5hmC.



Base	Sequence	BS Sequence	oxBS Sequence
C	C	T	T
5mC	C	C	C
5hmC	C	C	T

Figure 5-1: Overview of oxBS chemistry. Image courtesy of: Booth MJ, Branco MR, Ficz G, Oxley D, Kreuger F, Reik W, Balasubramanian S. *Science* 2012, **336**(6083): 934-937.

Global 5hmC levels are known to be lower than 5mC levels and more variable between different tissues (Globisch et al., 2010; Nestor et al., 2011). Here, a novel method is presented for genome-wide profiling of 5hmC by coupling oxBS chemistry and the Infinium 450K BeadChip (oxBS-450K). Alternative approaches also using oxBS or TET-assisted bisulphite (TAB) chemistries have been developed by Field et al. (Field et al., 2015) and Chopra et al. (Chopra et al., 2014), respectively. In order to maximise the probability of successful 5hmC detection, two human brain samples were selected for profiling as it is thought to have the highest 5hmC levels of any tissue: a commercial brain DNA sample from an unidentified region referred to as 'brain1' and a frontal cortex DNA sample, referred to as 'brain2'. In order to test the range of possible 5hmC detection, a whole blood sample was also profiled for comparison, as previous studies have shown blood to contain one of the lowest levels of 5hmC (Nestor et al., 2011). 5hmC was successfully measured in all three samples by comparison of BS- and oxBS-450K datasets and validation of the reported 5hmC signal is provided using two independent technologies: mass spectrometry quantification of 5hmC and oxBS-pyroseq assessment of selected CpG sites.

5.2.2 5hmC signal detection using the 450K BeadChip

5hmC detection was achieved by identifying differentially methylated CpG sites between the BS- and oxBS-treated replicates within each sample set. The beta (β) value associated with each probe on an array reflects the methylation level at that particular location on a scale of 0 to 1, where 0 is unmethylated and 1 is fully methylated. β values resulting from BS-only treatment represent the total methylation score (5mC and 5hmC) as both cytosine modifications are resistant to deamination with comparable efficiency, resulting in both modifications contributing to the overall methylation signal. In contrast, β values resulting from oxBS treatment represent only the 5mC level at the corresponding probe locations; 5hmC is selectively converted to 5fC (whilst 5mC remains unchanged) during the initial oxidation step and therefore becomes susceptible to deamination during subsequent BS conversion.

In order to assess the reproducibility of signal generated by the 450K array from oxBS-treated DNA, all samples were processed in duplicate in a trial of the oxBS-450K workflow, using 500ng input DNA per sample per conversion (either BS or oxBS) as recommended for traditional BS conversion of DNA for 450K analysis. Unsupervised hierarchical clustering of the resulting dataset shows a clear separation between blood, brain1 and brain2 and also between treatment options, with both BS-only and oxBS-treated replicates of each sample clustering in pairs (Figure 5-2).

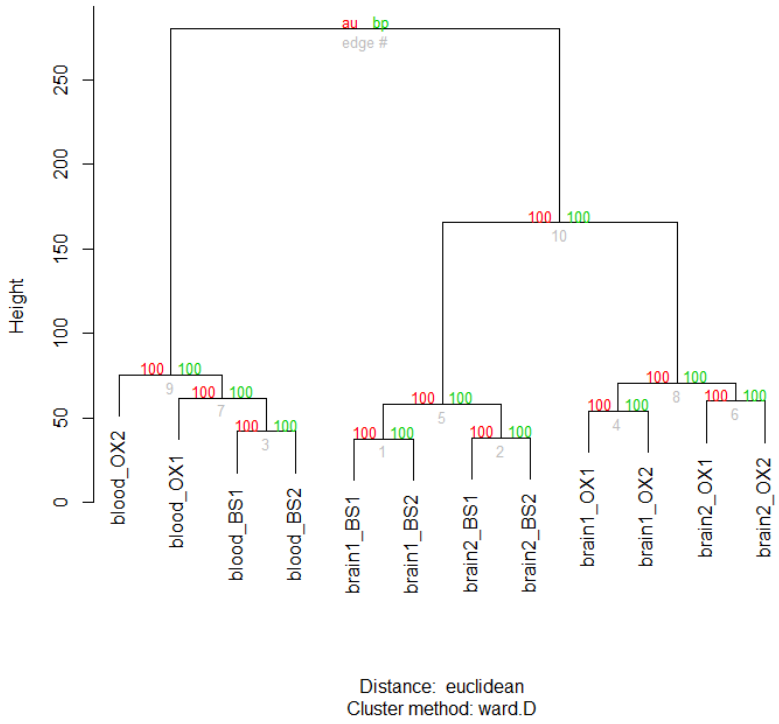


Figure 5-2: Unsupervised hierarchical clustering dendrogram of oxBS-450K samples. BS-treated samples are indicated by "BS" label; oxBS-treated samples by "OX". Numbers in red and green indicate approximately unbiased (au) and bootstrap (bp) probability values (%), respectively. Number of bootstrap resamples = 1000.

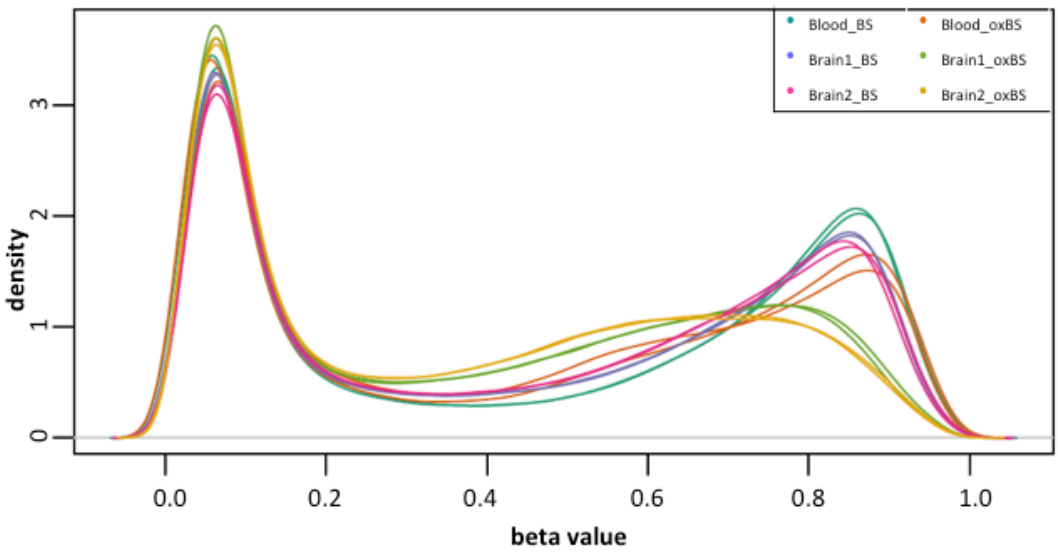


Figure 5-3: Kernel density plot of normalised β values generated by the oxBS-450K arrays. Beta values show a proper bimodal distribution for BS-only conversion samples, whilst a left-skewed methylated peak is seen in the oxBS conversion samples, corresponding to lower global methylation levels.

A kernel density plot of all β scores for each sample shows a clear reduction in overall methylation after oxBS treatment as compared to BS-only treatment for both brain1 and brain2 (and to a lesser extent, blood) as would be expected if 5hmC were present in the sampled DNA (Figure 5-3).

In order to identify potential 5hmC, normalised β values were used to calculate delta beta (Δβ) values for each probe by subtraction of the oxBS β value from the BS-only beta value. The Δβ score is a reflection of the 5hmC level at each particular probe location. Using this method, both positive and negative Δβ scores were calculated. Positive Δβ values represent potential sites of 5hmC; negative Δβ values, or 'hypermethylation' in the oxBS-treated sample, were not expected as the oxidation reaction is unidirectional and thus are likely to reflect background noise generated by this method. Visualisation of the Δβ distribution across all probes for the brain1 sample reveals many more positive Δβ scores than negative, as expected (Figure 5-4).

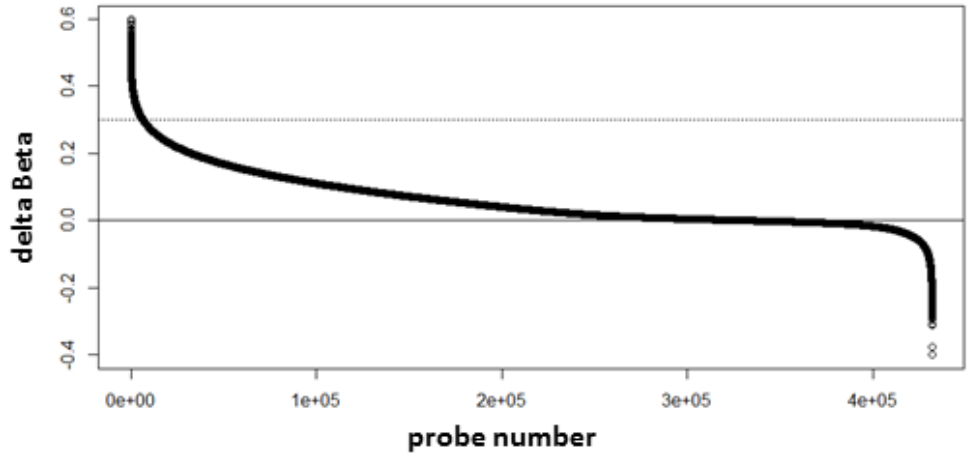


Figure 5-4: All normalised Δβ values for the brain1 sample. Calculated Δβvalues range from 0.6 to -0.4, with the vast majority (>95%) falling above zero, indicating potential sites of 5hmC.

Less than 3% of all negative Δβ scores correspond to a 5hmC level of more than 10%, however over 30% of all positive Δβ scores correspond to a 5hmC level over 10%. As two replicates per condition (i.e. BS-only & oxBS) were generated for each sample, significance analysis of the Δβ scores could be conducted by computing a t-statistic for each probe in the array dataset using the ChAMP analysis package in *R*

(Morris et al., 2014). This analysis pipeline allows a p-value to be assigned for each Δβ calculation, which may be either unadjusted or adjusted to control the false discovery rate using the Benjamini-Hochberg (BH) method (Benjamini and Hochberg, 1995). The largest Δβ scores are associated with the most significant p-values and that significance falls away as the Δβ values approach zero (Figure 5-5). Frequency plots of all probes according to their associated p-values show that the majority of Δβ values above zero represent significant differences in methylation score ($p \leq 0.05$). In contrast, the p-values associated with negative Δβ scores are distributed across the entire range (0 to 1), with increasing numbers of probes associated with higher p-values (Figure 5-6). This is consistent with the idea that the negative Δβ scores represent false differences in methylation score between the paired BS-only and oxBS datasets.

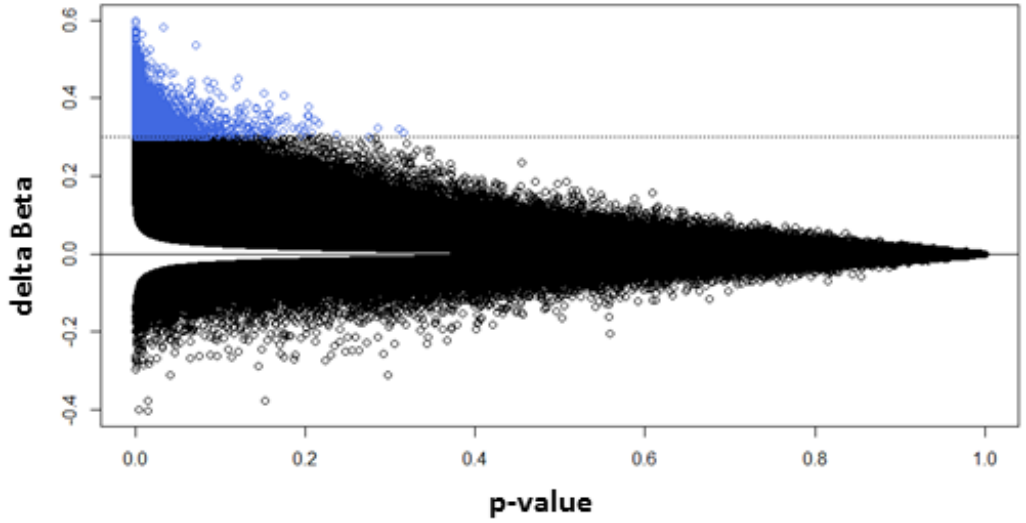


Figure 5-5: Scatter plot of all Δβ values against their associated p-value (unadjusted) for brain1. Δβ distribution against their associated p-value. Data points highlighted in blue indicate those probes called as hydroxymethylated ($\Delta\beta \geq 0.3$).

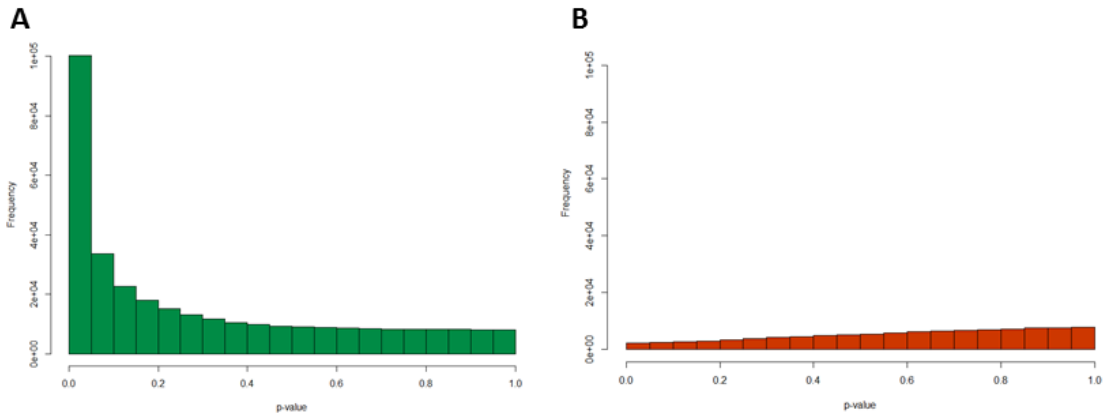


Figure 5-6: Frequency plots of all p-values associated with either positive Δβ (panel A) or negative Δβvalues (panel B). A) Frequency distribution of all probes associated with a $\Delta\beta > 0$, or potential sites of 5hmC. B) Frequency distribution of all probes associated with a $\Delta\beta < 0$, i.e. false positives.

There are various ways to define an appropriate Δβ threshold to identify a probe-set of significantly hydroxymethylated cytosines, including shrinkage estimation of variances but this approach requires each sample to be run in multiple (3-4) replicates resulting in extra costs [10]. As the amount of available genomic DNA is often limiting, the strategy employed here restricted the replicate number per sample to two, however this was not sufficient for any of the calculated Δβ scores to reach statistical significance (BH-adjusted p-value <0.05). A conservative threshold of $\Delta\beta \geq 0.3$ (i.e. minimum 30% 5hmC) was therefore applied alongside an unadjusted p-value cut-off of 0.05 for identifying hydroxymethylated probe locations, followed by validation. Using this approach, 6,578 and 7,692 probes were identified in the brain1 and brain2 samples, respectively, in contrast to just 801 probes in whole blood DNA. Inclusion of a third technical replicate of the brain1 sample, however, increased the statistical power of the dataset by allowing the Δβ calculations to reach statistical significance. A BH-adjusted p-value threshold of 0.05 was now suitable to call 5hmC, which dramatically increased the probe number because any significant Δβ values in the range of 0-30% 5hmC could now be identified alongside the more highly hydroxymethylated probes, resulting in the detection of 64,720 significant probes in total. Of these, only 117 sites were associated with negative (but nevertheless significant) Δβ scores, corresponding to a false discovery rate of less than 0.002 (Figure 5-7, panel A). 5hmC levels (i.e. all significant, positive Δβ scores)

fell within the range of 4-59%, with a mean level of 18% 5hmC; all negative Δβ scores (regardless of significance) corresponded to less than 20%.

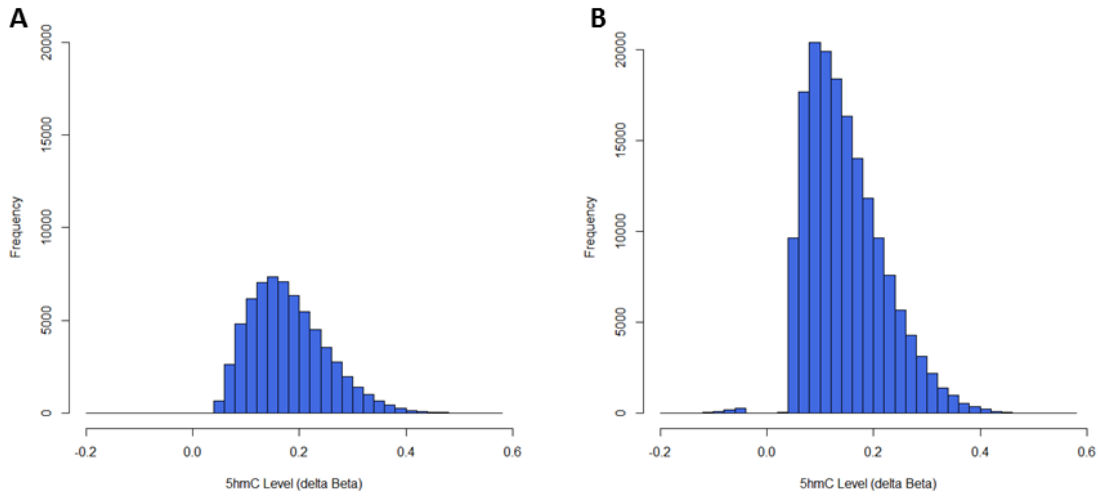


Figure 5-7: Frequency plots of Δβ scores for all hydroxymethylated probe sites identified using either two (panel A) or three replicate datasets for brain1 DNA. A) Frequency distribution of 5hmC levels, expressed as a Δβ score, across all 64,720 5hmC sites in the brain1 sample processed in triplicate with the initial protocol. B) Frequency distribution of 5hmC levels, expressed as a Δβ score, across all 165,495 5hmC sites in the brain1 sample, processed in duplicate with the optimised protocol. Both protocols detect similar Δβ ranges (5-60%).

5.2.3 5hmC distribution in human blood and brain samples

The distribution of 5hmC within different genomic features was investigated using probe sets of significant 5hmC sites compared to the overall distribution of probes present on the array. The following categories were used for classification of each probe: 3'UTR region, 5'UTR region, 1st exon, 200bp window upstream of the TSS (TSS200), 1500bp window upstream of the TSS (TSS1500), intergenic region or gene body. A permutation test was performed to calculate percentage enrichment for 5hmC within each feature compared to 1,000 matched, random 450K probe sets. 5hmC was found to be significantly enriched in the gene body in all three DNA samples (Figure 5-8). This is consistent with previous studies concerning the genomic distribution of 5hmC (Chopra et al., 2014; Szulwach et al., 2011). Conversely, 5hmC was found to be significantly depleted from promoter regions (defined as within 1500bp upstream of TSS) and from CpG islands (Figure 5-9), which agrees with several published studies that describe an inverse correlation

between 5hmC and CpG density at the promoter (Booth et al., 2012; Chopra et al., 2014; Yu et al., 2012a).

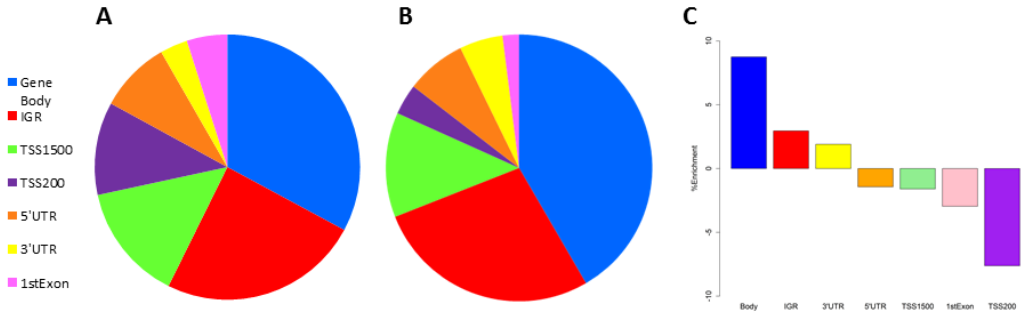


Figure 5-8: 5hmC enrichment within different genomic regions for brain1 sample. A) Distribution of all probes after filtering (432,056 total) on 450K BeadChip. B) Distribution of hydroxymethylated probes (6,578 total) in brain1 sample. C) Enrichment of 5hmC probes according to various genomic regions, compared to random 450K probe sets. Significant enrichment is seen in the gene body, and to a lesser extent 1500bp upstream of the TSS. Conversely, significant depletion is seen in the proximal promoter region (TSS200) and 1st exon.

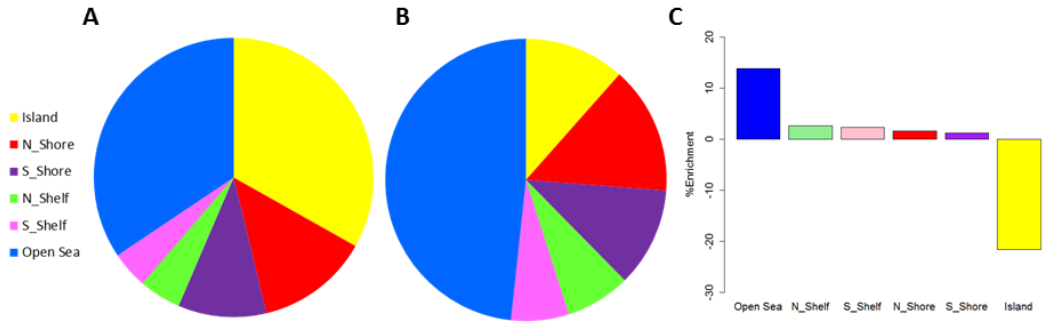


Figure 5-9: 5hmC enrichment within different CpG contexts for brain1 sample. A) Distribution of all probes after filtering (432,056 total) on 450K BeadChip according to CpG context. B) Distribution of hydroxymethylated probes (6,578 total) in brain1 sample according to CpG context. C) Relative enrichment of 5hmC probes in differing CpG contexts. Significant enrichment of 5hmC is seen in 'open sea' regions of low CpG density, although the most striking observation is the depletion of 5hmC at CpG islands.

5.2.4 Optimisation of sample processing for oxBS-450K analysis

An assessment of the reproducibility between oxBS-450K replicates revealed far higher variation compared to traditional bisulphite 450K array processing. The correlations between oxBS-converted replicates were the lowest, with Pearson correlation coefficients (r) ranging between 0.95-0.97 across the three samples. BS-converted replicates correlated slightly better, but their correlation coefficients were still lower than the 0.99 expected for 450K datasets (Table 5-1).

In order to improve the quality and reproducibility of signal generated from oxBS-450K arrays, several adjustments were made to the protocol in order to maximise the recovery of high-quality converted DNA at the end of the workflow. First, input DNA was increased from 500ng per conversion to 2µg to compensate for the additional centrifugation and purification steps present in the oxBS-450K protocol that reduce overall DNA yield as compared to standard BS conversion. An initial sonication step was included to shear high molecular weight gDNA to ≤10kb fragments in order to increase DNA recovery from the BioRad® P6 spin columns used in subsequent steps of the oxBS protocol and all DNA clean-up steps were performed with Agencourt Ampure XP beads instead of column-based purification methods, which minimised DNA loss. In addition, 7µl of eluted DNA was used for processing on the 450K BeadChip instead of 4µl DNA, as recommended by the manufacturer's protocol. This ensures a maximum amount of DNA is hybridised to each array on the BeadChip and compensates for the lower DNA yield of the TrueMethyl protocol in comparison to the EZ DNA Methylation kit (Zymo Research) used for standard BS conversion of DNA.

While the low-input oxBS-450K protocol only requires 1µg starting DNA, the high-input protocol requires 4µg DNA but results in improved sensitivity and reproducibility. Using the same detection p-value threshold of 0.01 as for the low-input protocol to filter probes, the technical replicates showed an improved correlation of $r > 0.99$ compared to $r > 0.95$ for the low-input protocol and comparable to current 'gold-standard' BS conversion protocols. Comparison of BS-only and oxBS datasets using just two replicates per sample was sufficient to call 165,495 5hmC sites (BH-adjusted p-value <0.05) with just 627 false positive probes

associated with a negative Δβ score (Figure 5-7, panel B). 5hmC levels were detected in the range of 4-57%, with a mean 5hmC level of 15%. In contrast, negative Δβ values all fell below 20%.

Analysis of the significant but negative Δβ probes between the low- and high-input protocols for the brain1 sample suggests that the locations of these negative sites are random: 117 sites were called by the low-input (three replicate) method compared to 627 with the high-input method, yet only eight negative Δβ sites were found in common between the two (the increase in negative probe number is a reflection of an overall increase in the number of significant probes identified using the high-input method). Despite showing statistically significant β value differences, these pseudo false positives cannot reflect true differences in methylation level at these probe sites and should be filtered out before further downstream data analysis.

Comparison of the 5hmC-containing probes identified by the different protocols showed that overlap is high, supporting the conclusion that these are sites of genuine 5hmC. 83.6% of the 5hmC-containing sites initially called by the low-input, two-replicate method remained significantly hydroxymethylated with the addition of a third replicate (5,498 of 6,578). This figure increased to 94.2% when comparison was made to the high-input, two-replicate dataset. Moreover, only 3.5% of the 5hmC probes called by the low-input, two-replicate method failed to reach significance when either replicate number or DNA input is increased, suggesting that the false positive rate of the low-input, two-replicate method is fairly low (Figure 5-10).

In summary, the use of the high-input oxBS-450K protocol on just two replicates resulted in detection of 5hmC at over twice the number of probes compared to the low-input protocol using three replicates. However, both the mean and range of 5hmC levels detected is very similar in both cases, suggesting that using lower amounts of input DNA does not necessarily affect the quality of the array data produced, but does severely limit the number of probes whose Δβ values reach statistical significance.

Sample type	Correlation coefficient (r)
Whole blood (BS)	0.982
Whole blood (oxBS)	0.950
Brain (BS)	0.985
Brain (oxBS)	0.963
Frontal cortex (BS)	0.984
Frontal cortex (oxBS)	0.951

Table 5-1: Pearson product-moment correlation coefficients (r) between replicates of each sample type processed using the low-input oxBS-450K protocol. Correlations are based on comparison of all normalised beta scores between each replicate pair.

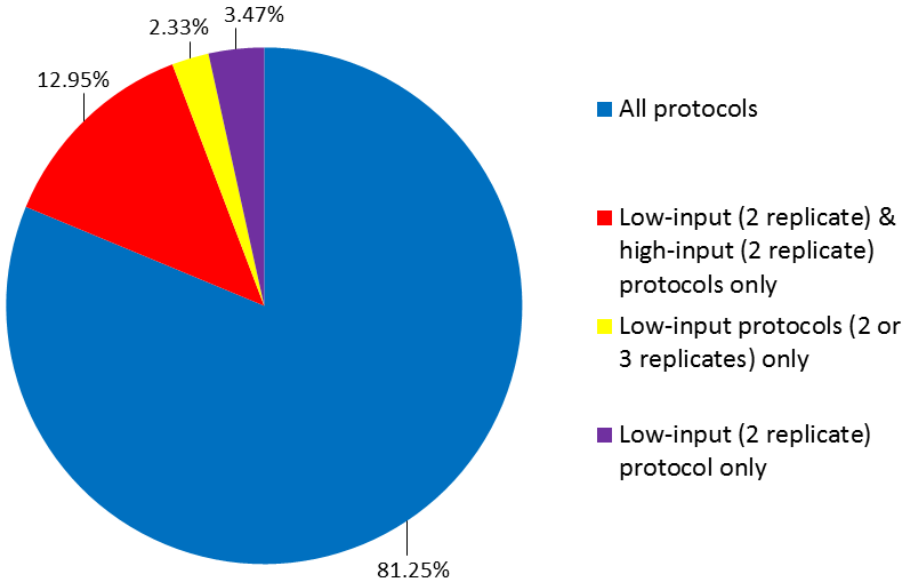


Figure 5-10: Hydroxymethylated probe overlap between the low- and high-input oxBS-450K protocols. Overlap is shown for the 6,578 probes called as hydroxymethylated using the low-input (two-replicate) protocol. Of these, 81.3% remained significantly hydroxymethylated both when the replicate number was increased and when the experiment was repeated using the high-input protocol. Only 3.5% of these probes were unable to reach significance when either replicate number was increased or when higher amounts of input DNA were used.

5.2.5 Validation of oxBS-450K methodology

Two methods, liquid chromatography-mass spectrometry (LC-MS) and oxBS conversion coupled to pyrosequencing (oxBS-pyroseq), were used to validate the 5hmC signal identified with oxBS-450K. On the global level, quantitative LC-MS was used to confirm that each of the profiled samples did indeed contain appreciable amounts of 5hmC, suitable for detection with oxBS-450K. Taking the average readings of two replicates, 5hmC levels of both brain samples were confirmed to be highly similar, at around 1% of total cytosine bases (0.92% in brain1 and 1.11% in brain2), whilst the whole blood sample contained only low levels of 5hmC, at 0.04% of total cytosine as expected. In contrast, 5mC levels in all three samples were broadly consistent at around 4% of total cytosine. The overall pattern of total 5hmC content per sample correlated well with the observed 5hmC signal from the 450K array data, where a ten-fold higher number of probes with significant 5hmC was observed in the brain as compared to whole blood (Figure 5-11).

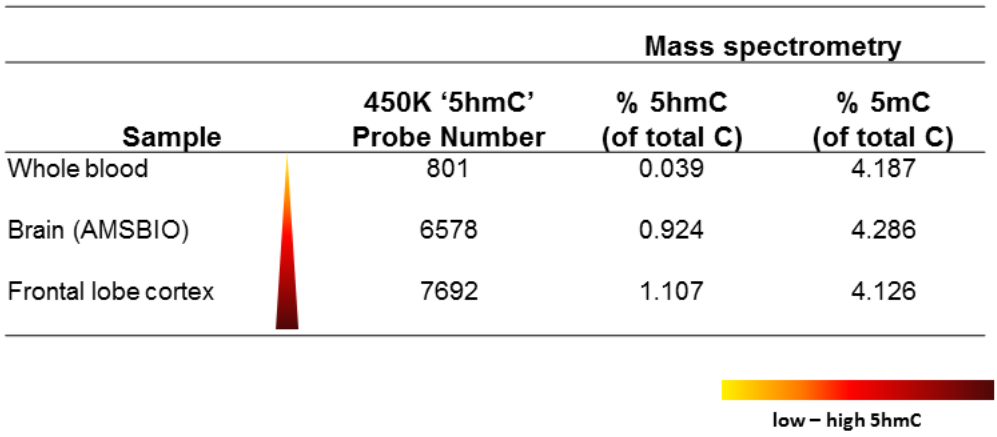


Figure 5-11: Quantification of global 5mC and 5hmC levels by mass spectrometry in three genomic DNA samples. The fraction of 5mC appears consistent across all samples, in contrast to the 5hmC content, where nearly 30-fold higher levels are seen in the two brain samples as compared to whole blood. This pattern of increased levels of 5hmC in brain is reflected in the 450K dataset (low-input protocol results are shown), where the numbers of CpG sites containing the highest levels of 5hmC (≥30%) are around ten-fold higher in brain versus blood. 5hmC percentage was confirmed by quantitative LC-MS to be 0.039% for whole blood and 0.924% and 1.107% for the two brain samples, respectively. In comparison, the corresponding 5mC levels were ~4% in all three samples.

On the single cytosine level, quantitative oxBS-pyroseq was used to validate selected 5hmC sites identified with oxBS-450K in the brain1 sample. The hydroxymethylation status of six individual CpG sites across three genomic regions

was confirmed using this assay (Figure 5-12), with 5hmC levels differing by less than 10% between the two methods. In addition, these validation results suggest that there is little difference in the accuracy of 5hmC calling depending on the 450K protocol used, as both low-input and high-input methods generated 5hmC levels within 10% of the pyrosequencing readout.

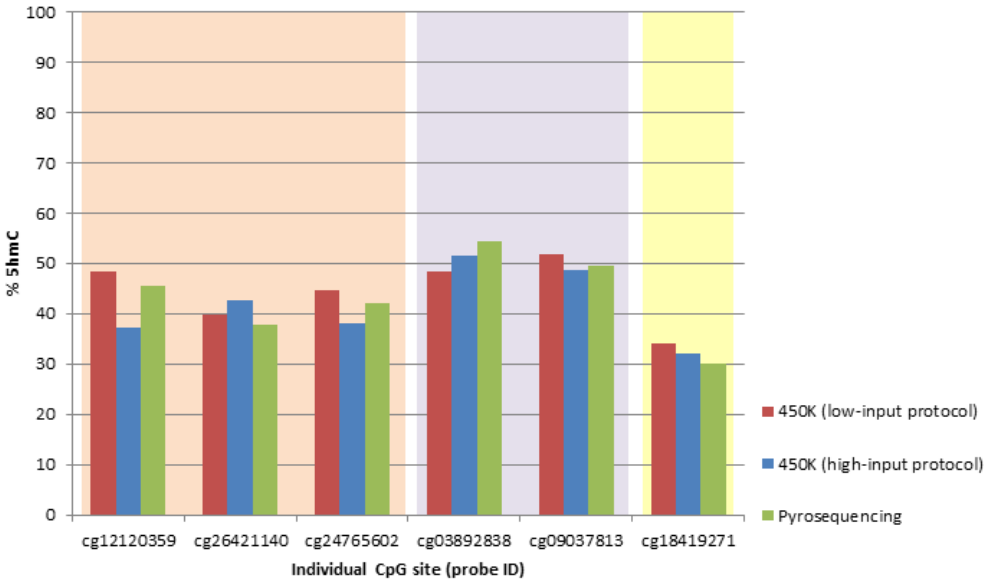


Figure 5-12: Pyrosequencing validation of 5hmC levels (%) at six individual CpG sites. 5hmC levels at six individual CpG sites distributed across three distinct genomic regions (indicated by orange, purple and yellow shading, respectively) were validated in the brain1 sample. Results are plotted as % 5hmC at each CpG locus as measured by pyrosequencing, low-input 450K and high-input 450K assays.

5.2.6 Discussion

The data presented in this chapter shows that oxBS chemistry can be combined with the 450K DNA methylation microarray to reliably detect 5hmC and 5mC in the human genome. The oxBS-450K method results in highly reproducible technical replicates, comparable to current 'gold-standard' BS conversion kits. 5hmC can be detected at statistically significant levels at over 30% of all CpG sites interrogated on the array, with a mean 5hmC level of 15% in a human brain sample. Whilst the use of the optimised protocol offers an improvement in the correlation of technical replicates and allows a higher proportion of 5hmC sites to be called with statistical significance, the amount of input DNA required may prove limiting for many users. However, validation using oxBS-pyroseq suggests that accurate 5hmC levels, albeit

at fewer sites, can nevertheless be called using a lower input of 1µg DNA for oxBS-450K analysis. Mass spectrometry measurements of global 5hmC and 5mC levels in human brain tissue agreed with previous findings showing 5hmC to be a significant contributor to the modified cytosine content of neural tissue types (Nestor et al., 2011). 5hmC made up approximately 20% of modified cytosine in the two assessed brain samples, highlighting how significant the overestimation of 5mC levels could be if relying on traditional BS-based methods of methylation profiling for the analysis of particular tissue types.

5.3 Development of an *in vitro* culture method for targeted demethylation studies

5.3.1 Introduction

Aberrant methylation phenotypes are a major hallmark of disease, particularly cancer (Feinberg and Vogelstein, 1983). Both global and site-specific hypo- and hypermethylation events have been observed during cancer progression (Feinberg, 2007); promoter hypermethylation has been frequently associated with tumour suppressor gene silencing events, such as p16 inactivation, in a wide array of human cancers, including colorectal and non-small-cell lung carcinoma (Herman et al., 1995; Merlo et al., 1995).

More recently, altered 5hmC patterns have also been associated with cancer phenotypes; reduced global 5hmC levels have been measured in various diseased tissues compared to controls, including liver, brain, lung, skin and pancreatic tissue (Jin et al., 2011; Lian et al., 2012; Liu et al., 2013; Yang et al., 2013). These findings support a hypothesis of 5hmC as an intermediary product of a DNA demethylation pathway that becomes deregulated during cancer progression, resulting in hypermethylation and associated hypo-hydroxymethylation in the affected tissue. DNMT1 is known to have a greatly reduced preference for 5hmC at a hemi-hydroxymethylated CpG site as compared to 5mC: DNMT1 activity on a 5hmC substrate is approximately 60-fold lower than its methylated counterpart, suggesting that 5mC oxidation may be a mechanism of inhibiting DNMT1 activity during DNA replication (Hashimoto et al., 2012; Valinluck et al., 2004). Similarly, the UHRF1 protein, a co-factor of DNMT1 essential for its methyltransferase activity, has a similarly reduced preference for hemi-hydroxymethylated CpG sites at a level comparable to its affinity for fully methylated or fully hydroxymethylated DNA sites. 5mC oxidation may be a method of ensuring passive DNA demethylation through blocking recognition of hydroxymethylated CpG loci by the DNA maintenance methylation machinery.

Whilst there is still debate on whether DNA methylation alterations represent a causal mechanism behind human cancers, there has been much interest in developing therapeutic drugs that reverse these epigenetic changes to normal levels in diseased

tissue (Singh et al., 2013; Yang et al., 2010). DNA methyltransferase (DNMT) inhibitors are a major class of epigenome-modifying drugs currently being developed for clinical use, of which two have already been granted approval by the US Food and Drug Administration (FDA) for the treatment of myelodysplastic (MDS) syndrome subtypes (Kaminskas et al., 2005; Kantarjian et al., 2006). 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine (DAC), also known by their trade names Vidaza and Dacogen respectively, are nucleoside analogues first synthesised in 1964 (Píškala and Šorm, 1964). 5-aza and DAC exert their demethylating activity through incorporation into DNA during replication and subsequent sequestration of DNMT1, the maintenance DNA methyltransferase, due to irreversible covalent binding of the enzyme to the analogues at hemi-methylated CpG sites. The DNMT enzymes are eventually degraded, leading to hypomethylation after successive rounds of DNA replication and thus restoring a passive DNA demethylation process.

Unlike DAC, 5-aza is also incorporated into RNA and disrupts protein synthesis through a variety of mechanisms, including inhibition of tRNA methyltransferases, histone methylase activity, rRNA processing and reduction in mRNA stability (Cihak et al., 1974; Li et al., 1970; Lu and Randerath, 1980; Momparler et al., 1976; Wada et al., 2005). Several studies have shown statistically significant changes in gene expression profiles following *in vitro* 5-aza or DAC treatment of various cell lines (Gius et al., 2004; Suzuki et al., 2002). In 2002, publication of one of the largest clinical trials investigating the treatment potential of 5-aza in 191 MDS patients showed a response rate of 60%, an increase in median survival rate and a significant improvement in quality of life for those patients receiving a low dose 5-aza treatment compared to those receiving only supportive care. These results made a considerable contribution to the FDA approval granted in 2005 for the use 5-aza to treat MDS.

However, the nonspecific mode of action of this class of DNMT inhibitors has caused concern over its potential use in the treatment of cancer. An *in vitro* study using pancreatic cancer cell lines showed an increase in the metastatic potential of four of the five cell lines tested after DAC treatments which was associated with the transcriptional activation of several matrix metalloproteinases critical for tumour cell invasion (Sato et al., 2003). Consequently, there is a real need to develop targeted

demethylating therapies to allow for tailored treatments to specific cancer subtypes and avoid the inadvertent activation of oncogenes in the process, which is a major concern with current globally acting, nonspecific demethylating agents (Feinberg, 2007; Gius et al., 2004).

One approach to achieve such targeted demethylation is through the use of an inactive, or caged, DNMT inhibitor whose subsequent activity can be induced by an external signal, in this instance UV light. Photocaged 5-azacytidine (caged-aza) comprises of two covalently linked photosensitive moieties at the N4 position of the pyrimidine ring of 5-azacytidine, which may be removed via photolysis upon exposure to light to produce the biologically active molecule. Photocaged 5-aza-2'-deoxycytidine (caged-DAC) was efficiently synthesised with just one photosensitive moiety attached at the same N4 position, affording the molecule fasting uncaging kinetics than the double-tagged 5-aza compound (Figure 5-13).

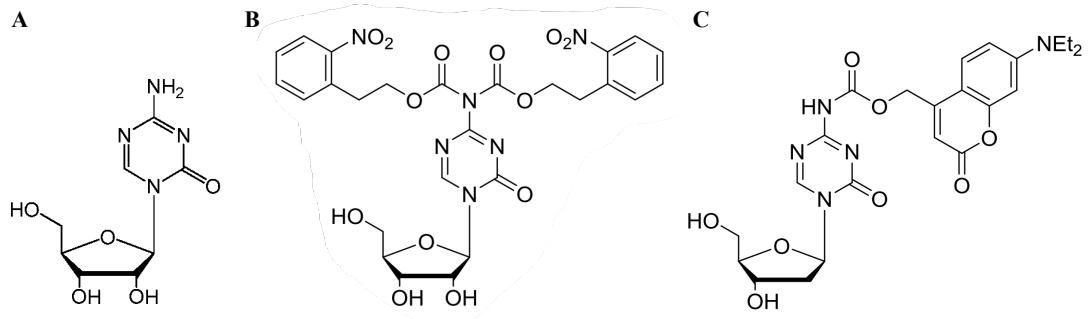


Figure 5-13: Molecular structure of the photocaged DNMT inhibitors. A) Unmodified 5-azacytidine. B) Photocaged 5-azacytidine ("caged-aza"). Two photosensitive moieties (2-nitrophenethyl carboxyl) are attached at the N4 position of the pyrimidine ring of 5-azacytidine. C) Photocaged 5-aza-2'-deoxycytidine ("caged-DAC"). One photosensitive moiety (7-diethylaminocoumarin-4-ylmethyl carboxyl) is attached at the N4 position of the pyrimidine ring of 5-aza-2'-deoxycytidine.

In this section, I outline a preliminary *in vitro* study into the use of the photocaged 5-aza and DAC molecules for targeted demethylation. A well-characterised osteosarcoma cell line (Saos-2), whose methylation profile had previously been established by the Beck lab, was chosen as the *in vitro* cell culture model for caged-aza treatment for its high proliferation rate and known hypermethylation phenotype, typical of many cancer cell lines. The stability of the photocaged molecules, their potential for temporally controlled activation through exposure to UV light and their

effects on the genome-wide DNA methylation profile of the Saos-2 cell line are discussed below.

5.3.2 Optimisation of 5-azacytidine *in vitro* treatment conditions

5.3.2.1 Concentration range

An initial titration experiment was performed using commercially available 5-aza to assess a suitable concentration range for use in subsequent experiments involving caged-aza that would result in a measurable decrease in methylation whilst minimising any potential cytotoxicity. Based on previous literature concerning 5-aza dosage (Hagemann et al., 2011), a concentration series in the range of 1-10μM was tested over a period of five days, with a total of three doses of 5-aza administered once every 24 hours to each treatment well. Five concentrations of 5-aza were tested in duplicate wells of 24-well culture plate: 1, 3, 5, 7 and 10 μM. The cells were harvested on day 5, 24 hours after the third and final dose of 5-aza.

On day 3, 24 hours after administering the first dose of 5-aza, a significant negative effect on cell growth could already be observed across all 5-aza treatment conditions. By day 5, the control wells had reached full confluence, whilst those undergoing 5-aza treatment had reached a maximum confluence of 50%, with increased inhibition of cell growth visible in the 5, 7 μM and 10μM treatment wells as compared to the lower doses of 1 and 3μM (Figure 5-13). This concentration-dependent increase in cell death was confirmed on day 5 after collecting all remaining cells for DNA extraction. The total amount of DNA recovered from the control wells was 24.4μg, compared to 13.2 μg for the 1μM treatment wells, 9.6μg for the 3μM treatment, 7.8μg for the 5μM treatment, 6.8μg for the 7μM treatment and 5.2μg for 10μM treatment, suggesting that the higher levels of cell death occurred in wells exposed to increasing concentrations of 5-aza.

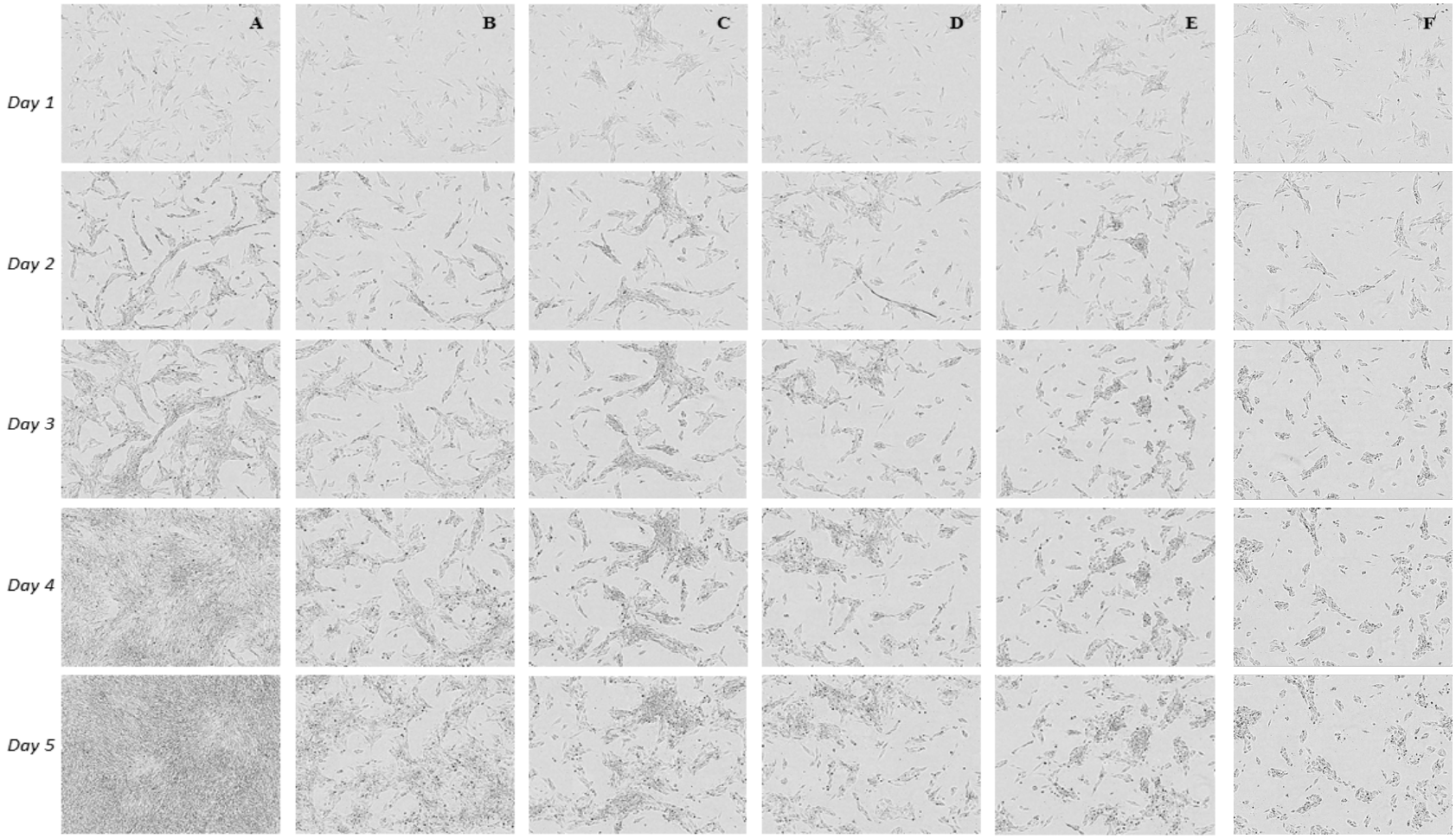


Figure 5-14: Concentration-dependent effect of 5-azacytidine treatments on cell growth. A) Control wells. B) 1µM 5-aza. C) 3µM 5-aza. D) 5µM 5-aza. E) 7µM 5-aza. F) 10µM 5-aza. First dose was administered on day 2; by day 3, a negative effect on cell growth is evident in all treatment wells (columns B-E) compared to control (column A). The control wells (column A) reached full confluence on day 5. In contrast, considerable cell death was observed in all wells receiving 5-aza, with higher concentrations (columns D and E) resulting in a more pronounced toxic effect.

Two factors were important when it came to determining an optimal concentration point for subsequent experiments testing the photocaged 5-aza and DAC molecules. First, it was critical to obtain sufficient DNA from each treatment condition to enable further downstream analysis of methylation levels either by an ELISA assay or Infinium 450K array; however, since typically only 500ng-1µg is needed per sample for either method, all 5-aza concentrations tested seemed reasonable. Second, it was important that the cells continued to undergo as many cell divisions as possible during the treatment time course, as 5-aza only exerts its demethylating effect through incorporation into the genome during DNA synthesis. Due to the replication-dependent nature of 5-aza bioactivity, the lowest concentration of 1µM 5-aza seemed the most appropriate as the toxicity of the drug is greatly reduced and only a moderate inhibition of cell division is observed under this treatment condition.

To confirm that a 1µM 5-aza treatment would be sufficient to generate a significant decrease in DNA methylation levels and thus successfully distinguish cells treated with 'inactive' photocaged 5-aza versus 'active' photolysed 5-aza compounds, global DNA methylation levels in the 5-aza treated samples were measured using a colorimetric ELISA assay.

A total of six samples were run in duplicate on an ELISA plate, along with a 5mC DNA standard curve. Whilst the 5mC standard curve did not show a linear correlation between absorbance and 5mC content as expected, a strong positive correlation was observed nevertheless (Figure 5-15). The highest absorbance reading, and hence DNA methylation level, observed amongst the six samples belonged to the control sample, i.e. untreated 'normal' Saos-2 cells, as expected. All samples subjected to 5-aza treatment showed a minimum 70% reduction in methylation levels compared to control, with an increasing 5-aza concentration from 1µM to 5µM correlating with a decreasing absorbance reading (0.0328 to 0.0114) and therefore global DNA methylation level. Interestingly, treatment with the two highest concentrations of 5-aza (7µM and 10µM) resulted in a reversal of this trend, with methylation levels higher than that observed for the 5µM-treated sample, albeit still at a fraction of the control sample's 5mC level (Table 5-2). This is likely due to

higher levels of cytotoxicity resulting from these treatments and consequently a greater restriction on cell growth and, crucially, DNA synthesis.

Based on these preliminary results, a 1µM concentration of 5-aza was deemed sufficient to cause a significant decrease in global DNA methylation levels whilst causing the least disruption to the proliferative capacity of the cells.

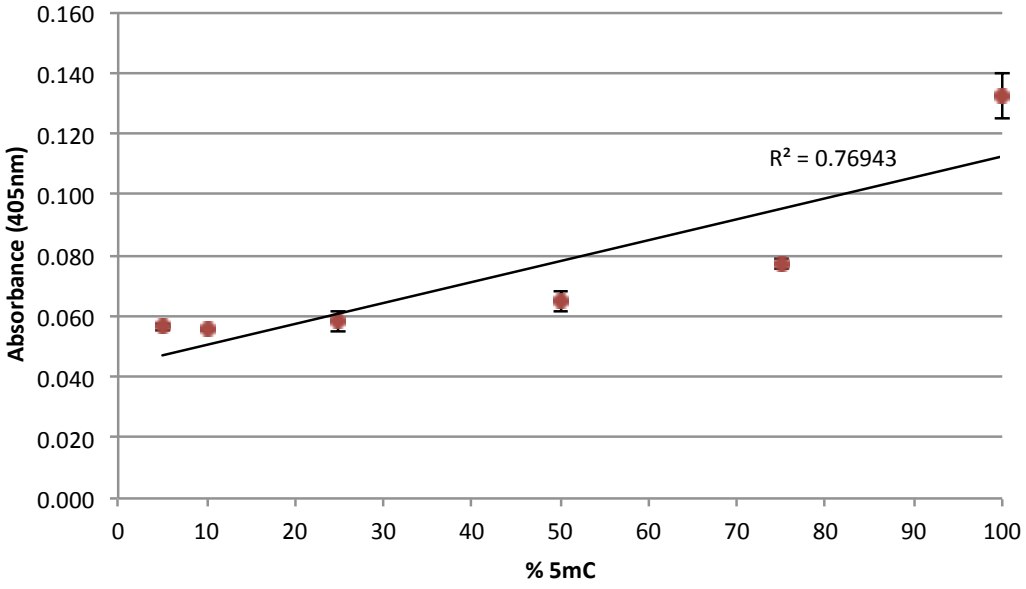


Figure 5-15: 5mC DNA standard curve generated using a colorimetric ELISA assay. Data points shown are the average of duplicate run for each sample. 5mC DNA standards were generated by serial dilution of a commercially available 100% methylated human DNA standard, creating samples containing 75%, 50%, 25%, 10% and 5% 5mC DNA.

Sample type	Raw absorbance reading at 405nm	Adjusted absorbance reading at 405nm
Control cells (no 5-aza treatment)	0.1666	0.1110
1µM 5-aza	0.0884	0.0328
3µM 5-aza	0.0751	0.0195
5µM 5-aza	0.0670	0.0114
7µM 5-aza	0.0701	0.0145
10µM 5-aza	0.0770	0.0214
Blank (DNA binding buffer only)	0.0556	0.0000

Table 5-2: 5mC DNA ELISA assay results for the 5-aza-treated Saos-2 cell samples. Absorbance readings shown are an average of duplicate samples for each treatment point. Adjusted absorbance reading = raw absorbance reading – blank measurement.

5.3.2.2 Solvent selection

Unlike 5-aza, the caged-aza compound is hydrophobic and needs to be dissolved in an alternative buffer in order to be administered in cell culture. Both dimethyl sulphoxide (DMSO) and acetonitrile (ACN) are two commonly used polar aprotic solvents for organic compounds, but are known to produce toxic side effects in living cells. Although exposure to these solvents could be minimised by using highly concentrated stocks of the caged-aza compound for the cell culture treatments, exposure cannot be entirely eliminated. A preliminary test was therefore performed to assess and compare the effect of pure DMSO and ACN on cell viability at the highest concentrations likely to be used in subsequent caged-aza experiments in order to select the most suitable dilution buffer.

In parallel to testing the effects of 5-aza at four different concentrations (0.5, 1.0, 1.5 and 2μM, dissolved in sterile water), both ACN and DMSO were also administered to cells in triplicate at a final concentration of 1.5μM, where one dose of the solvent was equivalent in volume to one 1.5μM dose of 5-aza. Both solvents had a surprisingly strong negative effect on cell growth compared to control, analogous to treatment with 0.5 to 1μM 5-aza (Figure 5-16A). However, the effect of ACN on cell growth was less severe, resulting in approximately 10% higher final cell confluence readings compared to DMSO-treated wells. This observed difference was amplified in a replicate culture plate, run in parallel to the treatment described above, which was left at room temperature under the cell culture hood for 1 hour after addition of each treatment before being returned to an incubator (set at 37°C, 5% CO_2). Under this condition, the 1.5μM ACN treatment resulted in an additional 25% confluence compared to DMSO (Figure 5-16B). However, the effect of the 1-hour exposure across all treatment types was minimal, suggesting that removal of the cells from the controlled environment of an incubator for 1 hour per day had little observable effect on cell viability. This was an important point to establish before proceeding with planned experiments to irradiate cell culture dishes containing the caged-aza compound. Irradiation would need to take place over the course of 1 hour outside of an incubator in order to produce sufficient quantities of the active compound via photolysis. As a result of this treatment course, ACN was chosen for use as the

solvent and caged-aza was stored at a stock concentration of 1mM at -20˚C and diluted further in ACN to 100µM just prior to use.

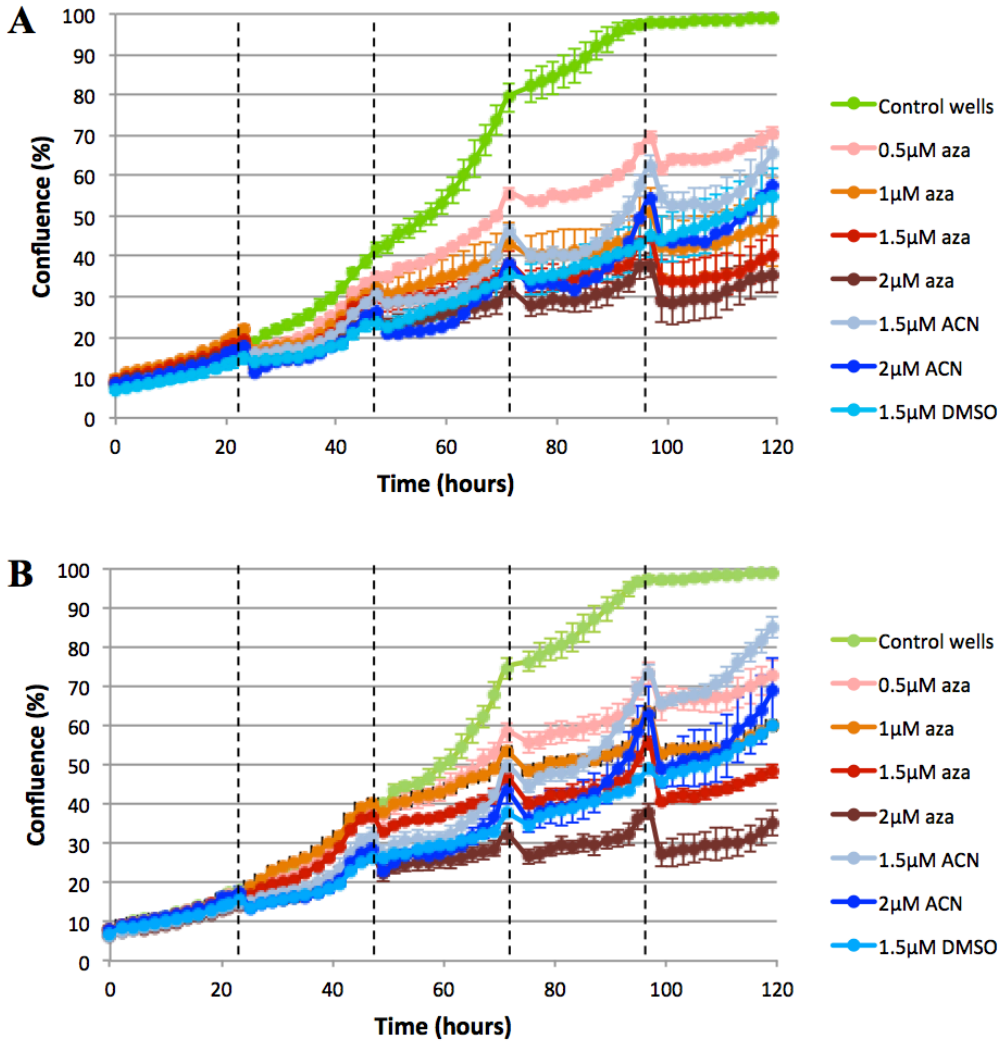


Figure 5-16: Growth curves demonstrating the effect of varying concentrations of 5-aza treatment against two alternative solvents, acetonitrile (ACN) and dimethyl sulphoxide (DMSO). (A) and (B) show the results from two replicate 24-well cell culture plates, both of which received identical treatments in parallel over the same time period. In addition, plate (B) was left at room temperature under the cell culture hood for 1 hour after addition of each treatment, in contrast to plate (A), which was immediately returned to an incubator (37°C, 5% CO_2). A concentration-dependent decrease in cell growth was observed in both plates for all 5-aza conditions tested. Control cells grew at near-identical rates in both plates, reaching a maximum confluence of 100% at the point of collection, suggesting that the additional time plate B spent outside of a controlled CO_2 incubator compared to plate A was well tolerated by the cells and did not have any negative effect on proliferation. NB: in both graphs above, time zero represents initial seeding of the culture plates.

5.3.3 Concentration-dependent effect of caged-aza on cell viability

The aim of these experiments was to establish whether the caged-aza molecule could be deprotected (i.e. activated) in a temporally controlled manner through irradiation of a cell culture dish after receiving a dose of the compound.

Prior to testing the demethylating potential of the photocaged 5-aza compound ("caged-aza") in cell culture, the following factors needed to be examined: first, the stability of the caged-aza molecule in culture media; second, the relative activity of the fully photolysed caged-aza compound compared to commercially available 5-aza; third, the feasibility of ensuring full deprotection of the caged-aza molecule through *in vitro* irradiation of a cell culture plate; and fourth, the possibility of toxic side effects resulting from the presence of the lysed phototag in the culture medium after deprotection and/or the hydrolytic breakdown of the caged-aza compound into harmful by-products.

5.3.3.1 Comparison of photolysed caged-aza versus 5-aza treatment *in vitro*

Solutions of both caged-aza and pre-irradiated (i.e. activated) caged-aza were freshly prepared and tested at three concentrations on the Saos-2 cell line over a period of four days, with one dose administered every 24 hours for a total of three doses before ending the time course on day 4, 24 hours after the third and final dose. The compounds were administered directly into the culture media to produce final concentrations of either 0.5, 1.0 or 1.5μM. Cell confluence was monitored throughout the course of treatment to assess the effect of the compounds on cell proliferation.

Both non-irradiated and pre-irradiated caged-aza were observed as having a deleterious effect on cell growth as compared to control cells receiving no treatment (Figure 5-17), although all three concentrations of the pre-irradiated treatments had a larger inhibitory effect on cell proliferation than any of the three non-irradiated caged-aza treatments, as would be expected of a biologically active 5-aza molecule. Furthermore, the lag in cell growth observed for the three non-irradiated, inactive compounds was comparable to that of the control cells receiving 1.5 μM doses of pure acetonitrile only, thereby mimicking the effect of the caged-aza solvent. It

seems reasonable to conclude that the inhibitory effect of the non-irradiated compound is likely to be due to the toxicity of acetonitrile, rather than the presence of active 5-aza in the culture media. Whilst the possibility of spontaneous photolysis in the culture media cannot be dismissed, the high rates of proliferation seen in all three non-irradiated treatments, resulting in >90% confluence across all three concentrations after the 4-day treatment, makes it unlikely that a significant proportion of the compound underwent any aberrant activation over the course of the treatment.

As expected, the pre-irradiated compound resulted in greater inhibition of cell proliferation at each concentration tested compared to both its non-irradiated equivalent and controls (Figure 5-17). The effect appeared more pronounced at higher concentrations, although considerable variation in confluence measurements across replicates means a definitive conclusion on the concentration-dependent nature of growth inhibition cannot be drawn. Interestingly, the pre-irradiated caged-aza did not replicate the growth inhibition to the same extent as previously observed in our treatments with 5-aza (Figure 5-16), with significantly slower growth observed over the treatment course of 5-aza from the time of the first dose (hour 24) to cell collection (hour 120), compared to the same time period for the pre-irradiated compound (hour 0 to hour 96 in Figure 5-17). This is likely due to incomplete photolysis of the caged-aza compound during irradiation, causing the concentration of active compound present in the irradiated solution to be lower than that of an equivalent concentration of 5-aza. This was confirmed by photolysis time course analysis of the irradiated caged-aza compound at 365nm, performed by collaborators, which found that 26% of the caged-aza compound was still present intact after 1 hour of irradiation, with the remaining 74% successfully uncaged to the active form.

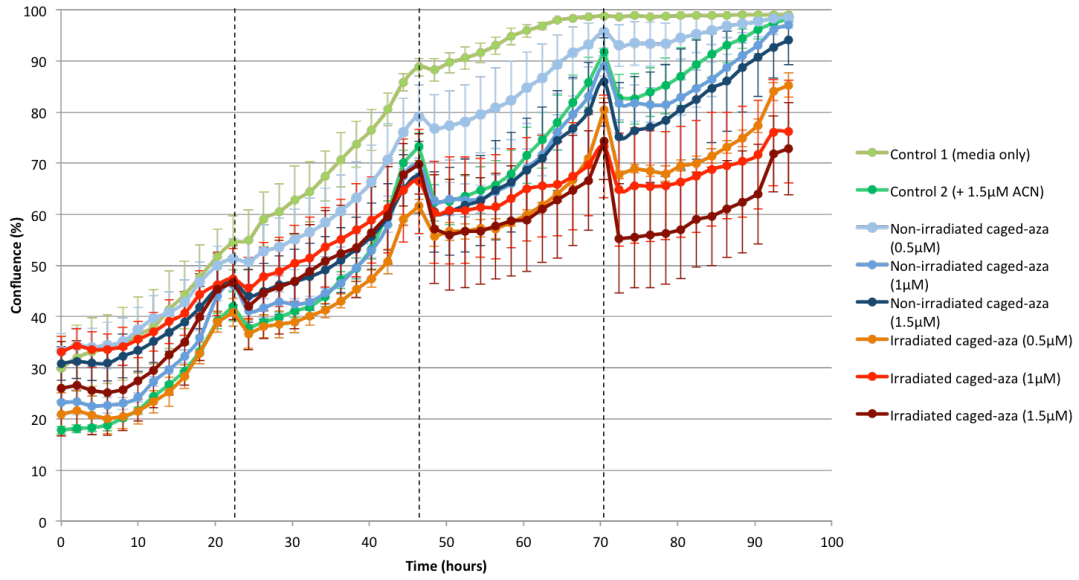


Figure 5-17: Growth curves of Saos-2 cells treated with varying concentrations of pre-irradiated or non-irradiated caged-aza. One dose was administered every 24 hours for a total of three doses, each indicated by dashed lines above. The cells were collected 24 hours after the final dose (at 96-hour time point, above). The non-irradiated caged-aza treatments had a less noticeable negative effect on cell viability than the pre-irradiated molecules, as would be expected for an inactive compound. All three concentrations of the inactive compound (0.5, 1.0 and 1.5µM) resulted in higher cell confluence readings at the end of the 4-day time course than the lowest concentration of active compound tested (0.5µM). Error bars represent the standard deviation in confluence measurements between three replicates for each treatment condition.

5.3.3.2 *In situ* irradiation of the caged-aza compound

The ability to successfully cleave the caged analogue after addition to cell culture was investigated as a first step towards achieving temporal control over the activity of this demethylating drug. Treatments were extended 1 day in comparison to previous experiments, in order to administer a total of four doses of caged-aza. Doses were added every 24 hours over a 5-day period at three different concentrations alongside appropriate controls, starting 24 hours after initial seeding of the culture plates. Treatments were run in parallel in two cell culture plates, one of which (plate B) was exposed to 1 hour of ultraviolet (UV) irradiation at 365nm immediately following the addition of each of the final three caged-aza doses. The control plate (A) remained protected from UV light throughout.

Previous work in the Howorka group established the absorption spectrum of the caged-aza molecule and identified that the maximum deprotection rate would occur through UV irradiation at 260nm. However, due to the known mutagenic effects of

UV exposure (Ichihashi et al., 2003; Ravanat et al., 2001), a longer wavelength (365nm) falling within the UVA spectrum of 320-400nm was chosen for the *in vitro* irradiation experiments in order to minimise DNA damage and overall cell toxicity. Although the measured absorption at 365nm was quite low (2%), after 60min of irradiation 74% of the caged-aza molecules had undergone photolysis. Assuming that the observed breakdown of caged-aza resulted in full deprotection of the compound (i.e. cleavage of both phototags), 1 hour of irradiation was therefore deemed sufficient to activate the caged-aza compound whilst minimising cell exposure to potentially damaging irradiation.

Growth curve profiles for cells grown in the control condition on both plates were almost identical, resulting in ~90% confluence by the end of the treatment course (hour 120), suggesting that UV irradiation on its own had no effect on the proliferative capacity of these cells.

As would be expected of a biologically inactive compound, caged-aza treatment in the absence of light exposure had little discernable effect on cell growth (Figure 5-18A). Growth curves for cells subjected to varying concentrations of caged-aza or pure acetonitrile had similar profiles to control cells exposed to full growth media only, suggesting the caged compound has little or no toxicity in its inactive form and is well tolerated by cells in culture. Treatment with the highest concentration of caged-aza, 1.5µM, did cause a small decrease in cell growth relative to other treatment conditions, as observed in previous treatments with the inactive caged-aza compound (Figure 5-17), likely due to either a small percentage of the compound spontaneously breaking down to produce uncaged 5-aza or to the increased volume of ACN present in each dose.

In contrast, caged-aza treatment in combination with UV irradiation produced a clear concentration-dependent reduction in cell proliferation (Figure 5-18B), with 1.5µM caged-aza resulting in almost complete cell death in the corresponding wells. Caged-aza treatment in the absence of UV irradiation had no discernable effect on cell growth, as shown by the tight clustering of growth profiles for all treatment conditions during the first 48 hours after initial seeding (time zero), during which both plates A and B were subjected to identical conditions, including the first

treatment dose at hour 24 and no UV exposure. However, the introduction of UV irradiation for plate B after the second dose at hour 48 resulted in immediate, visible reduction in cell growth rates for all three caged-aza concentrations.

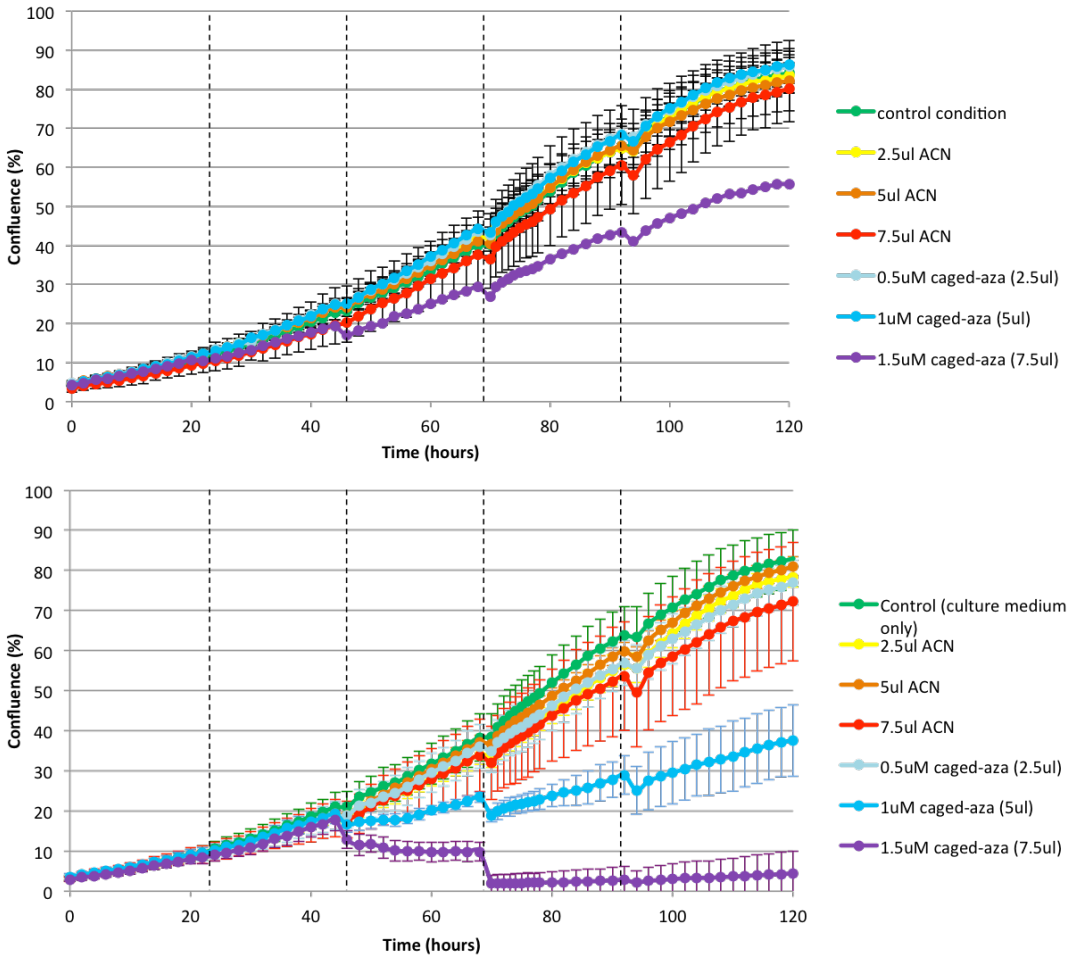


Figure 5-18: Growth curves demonstrating the effect of UV irradiation on the proliferation capacity of Saos-2 cells treated with caged-aza. Graphs A and B show the growth curve profiles of duplicate culture plates subjected to identical treatment conditions but differing light exposure. Both plates were administered four doses of caged-aza/ACN as appropriate at 24, 48, 72 and 96 hours post-seeding, indicated by the dashed vertical lines above. Plate A was protected from all UV exposure, whilst plate B was irradiated for 1 hour at 365nm a total of three times, directly following the addition of caged-aza/ACN doses at 48, 72 and 96 hours post-seeding. Irradiation of plate B was withheld after the first dose to mimic the conditions of control plate A.

5.3.3.3 Effect of photolysis by-products on cell viability

Saos-2 cells were treated with two concentrations of the isolated phototag (not conjugated to 5-aza) in parallel to the caged-aza treatments described previously in order to assess whether it produced any cytotoxicity of its own. Whilst the phototag was predicted to be pharmacologically inactive, it was important to confirm that the phototag would not produce any negative side effects once the caged-aza compound

underwent photolysis. Addition of the phototag over a 5-day treatment period at either 1 µM or 1.5µM concentrations had no appreciable effect on cell proliferation, even in combination with UV irradiation (Figure 5-19). All four photolysis by-product treatment conditions resulted in similar sigmoid growth curves, with all culture dishes reaching full confluence by the end of the time course. The rate of growth observed was comparable to treatment with caged-aza in the absence of UV (Figure 5-19; orange label), as expected. These results suggest that exposure to the lysed phototag does not result in cytotoxicity, or at least not at levels sufficient to induce senescence. The differential growth rate observed with caged-aza treatment (Figure 5-18) is therefore likely due to the activity of the uncaged cytidine analogue rather than the accumulation of phototags in the cell culture media.

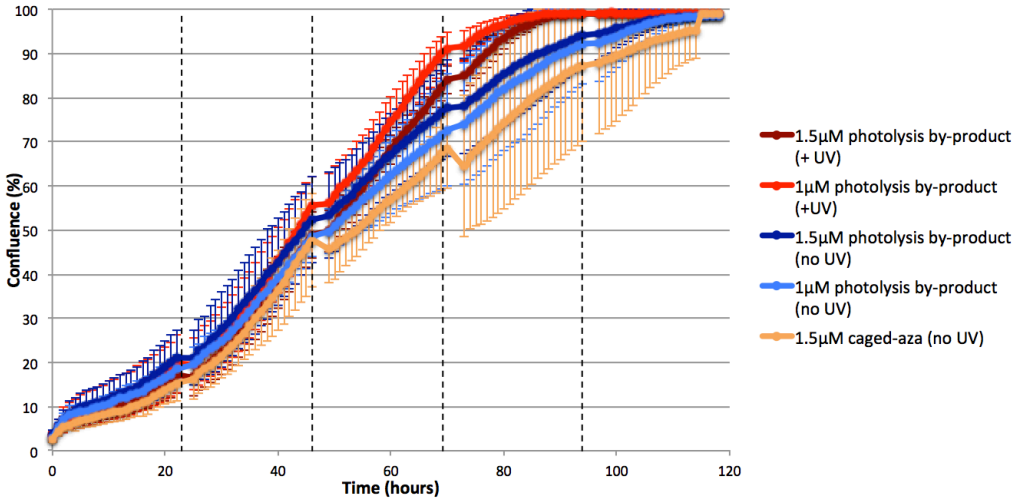


Figure 5-19: The effect of treatment with the photolysis by-product. Cells were treated with either 1 µM or 1.5µM solutions of the isolated photolysis by-product in both the presence and absence of UV irradiation at 365nm. Dashed lines (above) indicate dosage time points. A total of 4 doses of the phototag were administered every 24 hours in fresh media in two cell culture plates run in parallel. One plate was exposed to 1 hour of UV immediately after each dose; the second plate was protected from light throughout.

5.3.4 Genome-wide DNA methylation analysis of caged-aza treatment

Although the *in vitro* experiments described above showed promising results for the efficacy of the caged-aza compound, its demethylating activity needed to be confirmed before it could be considered as a viable alternative to traditional 5-aza. 12 DNA samples were subjected to genome-wide DNA methylation analysis using the Infinium 450K array. DNA was extracted from Saos-2 cells that had undergone the

following treatments: 5-aza, caged-aza and pre-irradiated caged-aza treatments at 0.5, 1.0 and 1.5µM. In addition, 3 control conditions were included: no treatments, UV irradiation only, and 1.5µM acetonitrile treatment combined with UV irradiation.

Data was processed using the ChAMP analysis pipeline in *R* (see Chapter 2, section 2.9.6). Normalised beta values for each probe on the array, representing the methylation level at each CpG site on a scale of 0 to 1, were obtained for each of the 12 samples. Comparison of the frequency distributions of beta values between samples highlights a clear reduction in DNA methylation following 5-aza treatment (aza; Figure 5-20). However, no such reduction in DNA methylation was seen for the caged-aza treatments combined with UV irradiation (cagedAzaUV; Figure 5-20). The density plots for the 'cagedAzaUV' samples were identical to the control and caged-aza samples, suggesting that this compound did not induce widespread DNA demethylation in the treated cells.

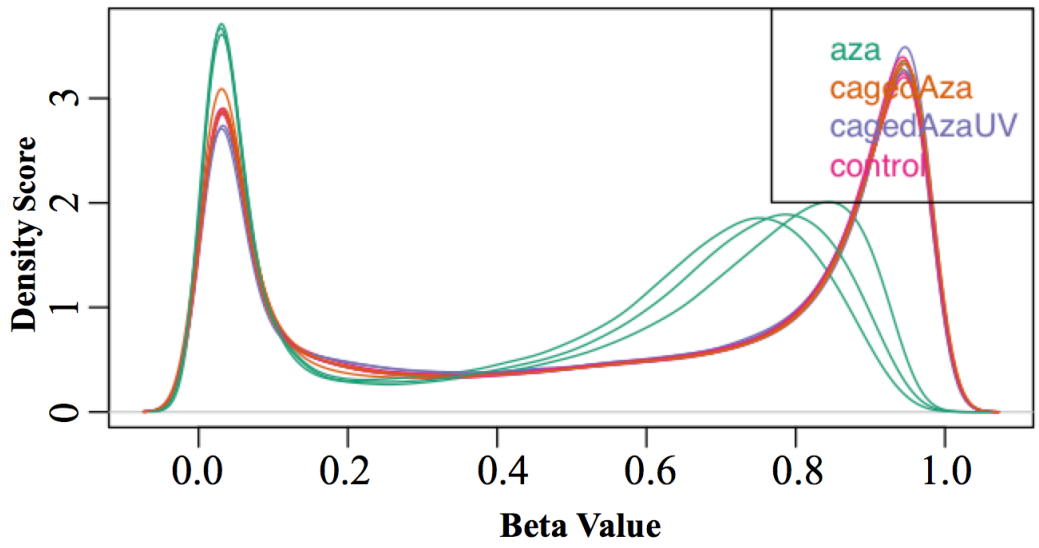


Figure 5-20: Kernel density plot of all normalised beta values. The frequency distribution of normalised beta values generated from the 450K array are plotted for all 12 samples and grouped according to treatment type. Samples subjected to 0.5, 1.0 and 1.5µM 5-aza are labelled as 'aza' (green). 'CagedAza' refers to samples subjected to caged-aza in the absence of UV (orange). 'CagedAzaUV' refers to samples subjected to caged-aza combined with UV irradiation (purple). 'Control' refers to the three control samples (pink).

Analysis of the degree of correlation between samples confirmed that the caged-aza treatments did not cause any alteration in the DNA methylation profile of the Saos-2 cells (Supplementary Figure 17). All three 'cagedAzaUV' samples had a correlation coefficient of 0.99 when compared to any of the three control samples. Similarly, comparison of 'cagedAza' samples to controls all resulted in correlation coefficients of 0.99 or above. In contrast, all three 'aza' samples showed much lower correlations when compared either to caged-aza samples or controls (r=0.78-0.79), suggesting that the 5-aza treatments did have a substantial effect on the DNA methylation profile of their respective cells, as had been expected. Treatment with 0.5μM 5-aza decreased median DNA methylation levels by approximately 10%, from 69% 5mC in the control cells to 62% in the treated cells. Increasing concentrations of 5-aza decreased 5mC levels still further, with 1μM and 1.5μM 5-aza treatments resulting in averages of 58% and 57% 5mC, respectively.

The lack of demethylating activity observed for the caged-aza treatments may be due to either insufficient uncaging or instability of the compound when dissolved in the cell culture media. High performance liquid chromatography (HPLC) analysis of the compound found that caged-aza was extremely prone to hydrolytic breakdown when in aqueous solution (mimicking cell culture conditions). Further HPLC analysis of the compound subjected to UV irradiation found that 52% remained unchanged after 1 hour of irradiation at 365nm and only 16% of the compound was fully uncaged. The remaining 32% had undergone cleavage of a single phototag to produce a mono-tagged 5-aza molecule.

5.3.5 Concentration-dependent effect of caged-DAC on cell viability

An alternative caged cytidine analogue containing a single phototag was synthesised in an attempt to overcome the slow uncaging kinetics and hydrolytic instability observed with the caged-aza compound. The phototag was exchanged for one with higher absorption at the desired wavelength of 365nm. In addition, DAC was chosen over 5-aza as the preferred demethylating agent to undergo modification due to its greater hydrolytic stability at 37ºC and lower cytotoxicity (Lin et al., 1981; Notari and DeYoung, 1975; Stresemann and Lyko, 2008). The resulting molecule, mono-DEACM-deoxyazacytidine (caged-DAC; Figure 5-13C), was then subjected to

HPLC analysis whilst undergoing irradiation at 365nm. 92% of the compound was converted to the biologically DAC after 1 hour, with only 0.5% of the original compound remaining intact.

Treatment of the Saos-2 cell line with varying concentrations of caged-DAC combined with UV irradiation at 365nm resulted in a strong concentration-dependent reduction in cell growth (Figure 5-21A), as previously observed with the caged-aza compound. All concentrations of caged-DAC resulted in a decline in the rate of cell growth, although the effect was much more pronounced at the higher concentrations of 4.5 and 13.5μM, with the latter resulting in almost complete cell death. These observations are consistent with the idea that the compound has undergone successful uncaging to its biologically active form, which can then incorporate into newly synthesised DNA. Secondary cytotoxic effects are often observed during 5-aza or DAC treatment, which are attributed to the decrease in DNMT1 levels rather than DNA demethylation activity (Jackson-Grusby et al., 1997; Jüttermann et al., 1994). The presence of increasing numbers of DNMT1-aza adducts facilitates mutagenesis at these CpG sites, leading an inhibition of DNA synthesis and cytotoxicity.

In contrast, treatment with caged-DAC in the absence of UV exposure caused only minor attenuation in growth over the 5-day treatment period, as would be expected from an inactive compound (Figure 5-21B). The minor reduction in cell growth observed for the highest caged-DAC concentrations is perhaps indicative of a low level of spontaneous uncaging of the compound. Further analysis of the stability of the compound in aqueous solution or an assessment of DNA methylation levels in these samples would be needed to definitively conclude whether spontaneous uncaging is occurring over the course of treatment.

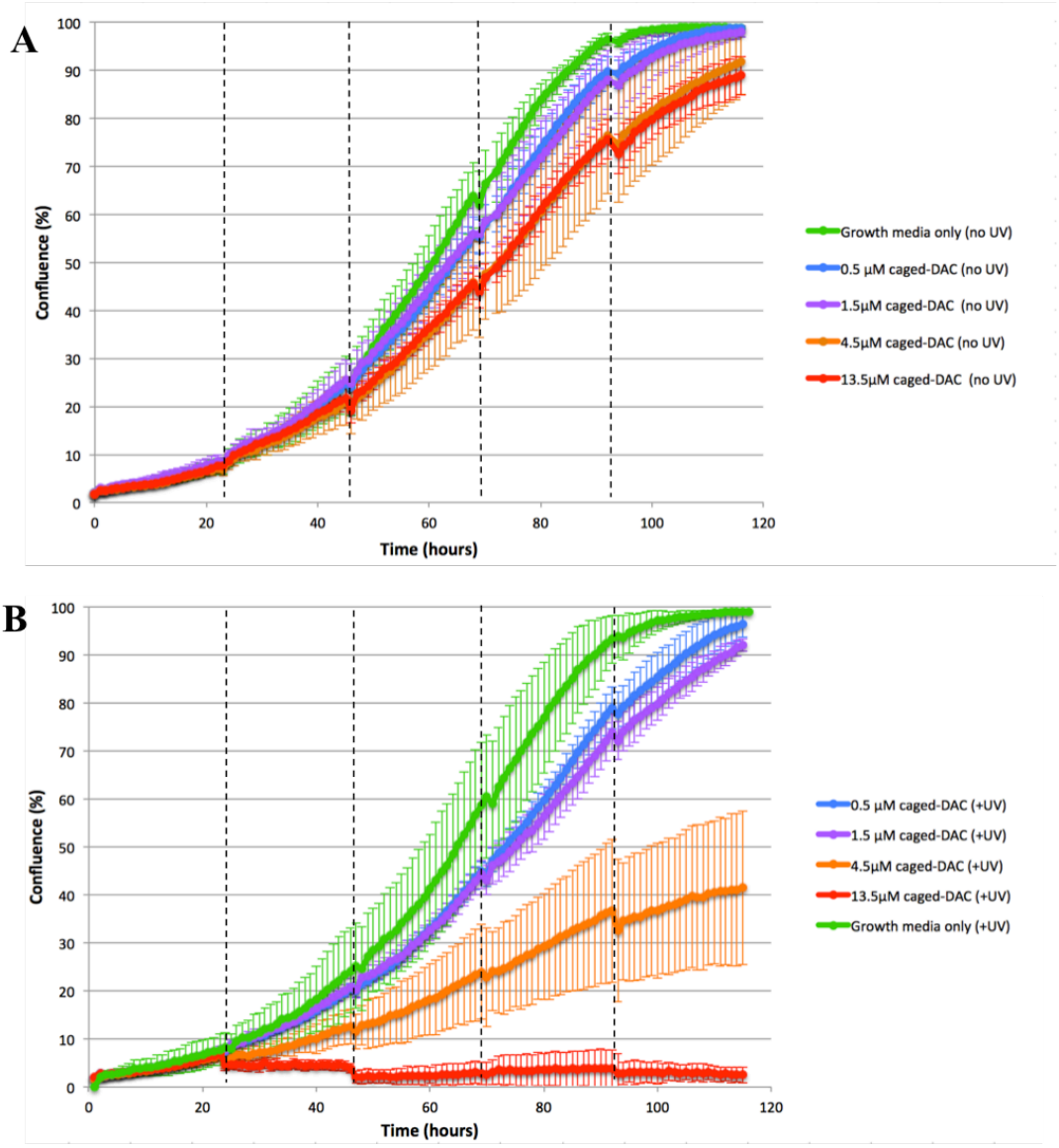


Figure 5-21: Negative concentration-dependent effect of caged-DAC treatment on cell growth. Saos-2 cells were subjected to 4 doses of either 0.5, 1.5, 4.5 or 13.5µM caged-DAC at 24-hour intervals in two parallel culture plates, one of which was exposed to 1 hour of UV irradiation following each dose. A) Cell confluence over time for all treatment conditions without UV exposure. B) Cell confluence over time all treatment conditions subjected to UV irradiation.

If full uncaging of the caged-DAC compound was successful during the given periods of UV irradiation, it would be expected to have a similar effect on cell growth as treatment with unmodified DAC. However, DAC treatment at all three concentrations tested (0.5, 1.5 and 4.5 µM) resulted in a larger inhibition of cell growth in comparison to the equivalent caged-DAC treatment (Figure 5-22B), suggesting that the caged-DAC compound does not undergo full uncaging under the irradiation conditions used here, perhaps due to partial UV absorption by the culture media or cells themselves, which would decrease the total amount of UV reaching

the compound. Alternatively, it is possible that the caged-DAC compound underwent partial hydrolysis to further inactive products.

Although UV exposure alone had no visible effect on cell proliferation ("growth media only" conditions, Figure 5-22A), UV in combination with DAC treatment appeared to have a much stronger inhibitory effect on cell growth compared to DAC treatment in the absence of irradiation. This was unexpected and may reflect the cells' inability to effectively counteract the increasing levels of DNA damage and cellular stress caused by the combined treatment.

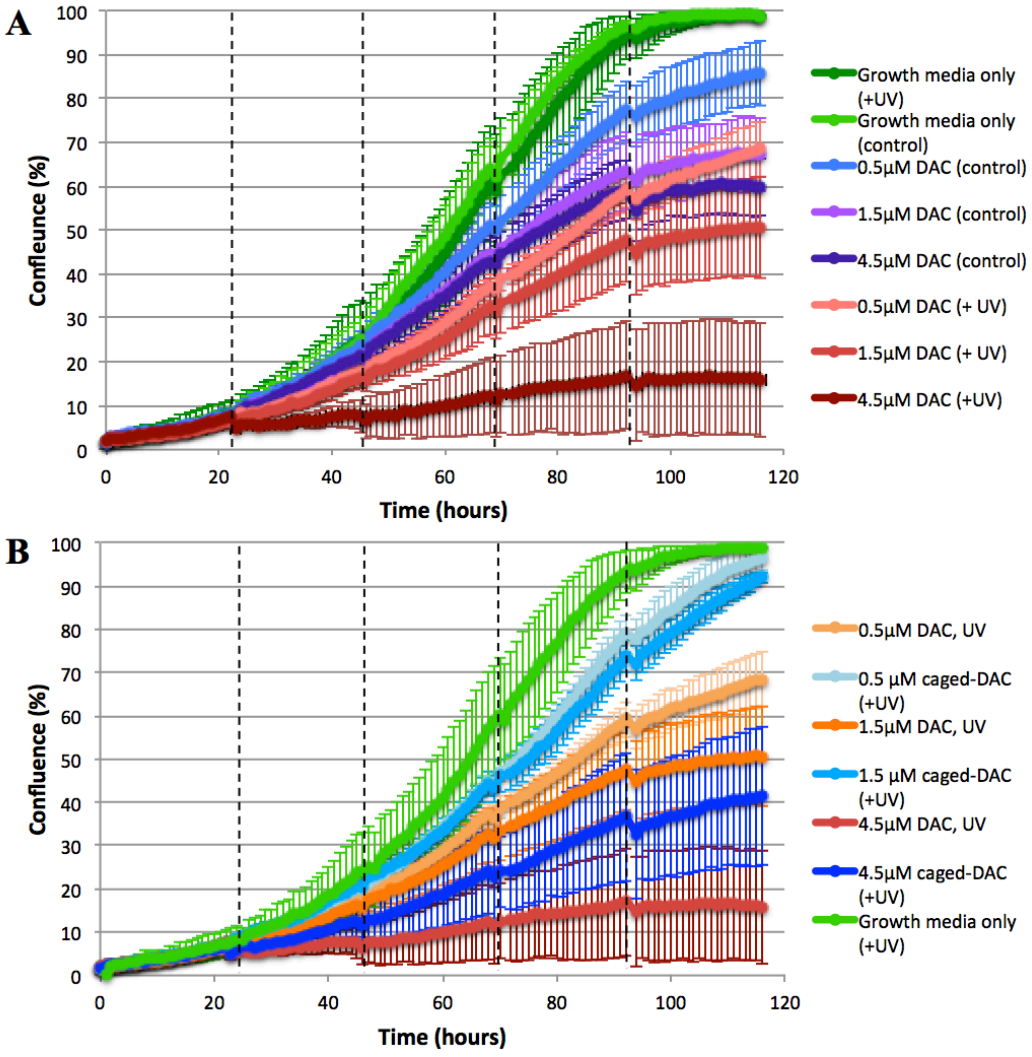


Figure 5-22: Effect of DAC treatment on cell growth. Three concentrations of both DAC and caged-DAC were administered to Sao2-cells in a total of four doses over five days. Treatments were given to duplicate cell culture plates, one of which was irradiated at 365nm for 1 hour following the addition of each treatment dose. A) Comparison of varying concentrations of DAC treatment with and without UV irradiation. B) Comparison of DAC versus caged-DAC treatment under UV irradiation.

5.3.6 Genome-wide DNA methylation analysis of caged-DAC treatment

The final aim of this chapter was to establish whether the caged-DAC molecule could produce a demethylating effect comparable to that of DAC after deprotection. DNA was extracted from Saos-2 cells treated with 1.5µM caged-DAC (both with and without UV irradiation at 365nm) and subjected to Infinium 450K BeadChip analysis. The raw data was processed using the ChAMP analysis pipeline in *R* alongside the previously generated 450K data for the caged-aza experiments (see section 5.3.4). The frequency distribution of normalised beta values across all probes shows a clear shift in the methylation profile of both caged-DAC samples towards a hypomethylated state (Figure 5-23). The 'CagedDAC_UV' sample shows the largest decrease in global methylation levels, indicating that successful uncaging of the caged-DAC molecule occurred *in vitro* to release the biologically active DAC. However, a similar but less pronounced effect was observed for the 'CagedDAC_control' sample that did not receive UV irradiation and was consequently expected to remain unaffected by the caged-DAC treatment. The median methylation score across all probes decreased from 68% in the control sample (Saos-2 cells under normal growth conditions) to 51% in the caged-DAC_control sample (compared to 36% in the caged-DAC_UV sample) suggesting that a significant proportion of the caged-DAC compound underwent spontaneous hydrolysis and was not stable in culture media.

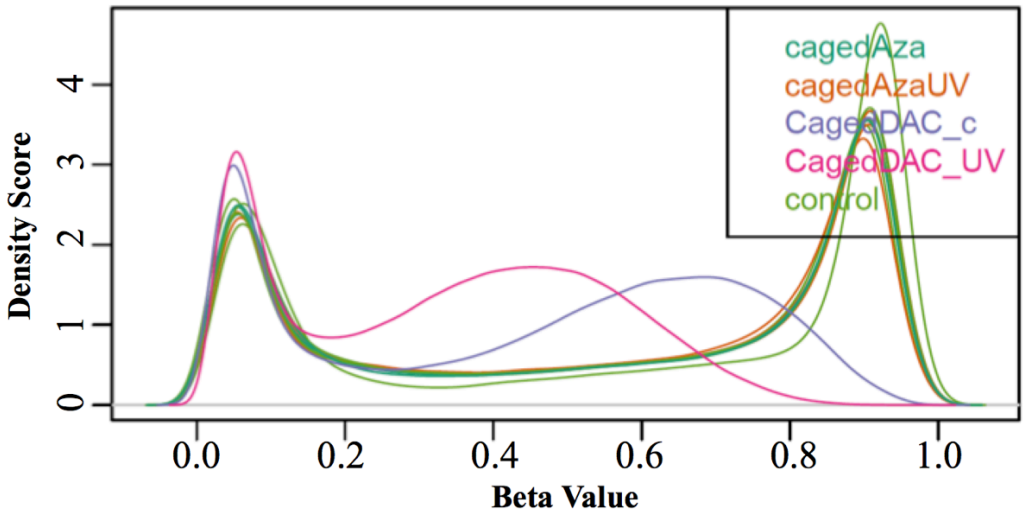


Figure 5-23: Kernel density plot of normalised beta values from the Infinium 450K array. Frequency distribution for all normalised beta values. 'CagedDAC_c' refers to treatment with caged-

DAC in the absence of UV; 'CagedDAC_UV' refers to treatment with caged-DAC combined with UV irradiation.

Visualisation of the average methylation level across all probes on the 450K array in both control and caged-DAC-treated samples (Figure 5-24B) shows that the demethylating activity of caged-DAC is not constricted to particular genomic regions but instead acts indiscriminately across the genome, producing an unbiased genome-wide hypomethylation phenotype.

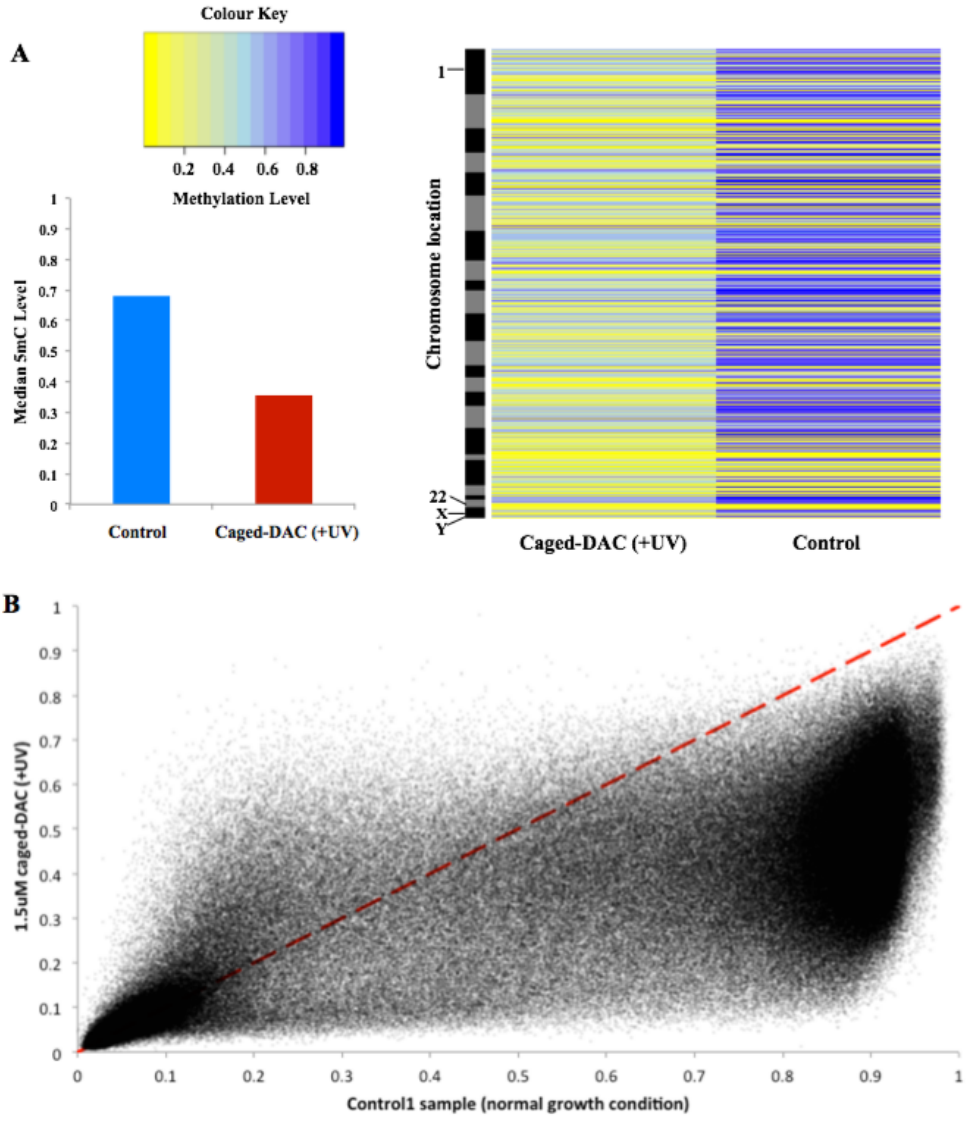


Figure 5-24: Effect of caged-DAC treatment on genome-wide DNA methylation profile of Saos-2 cells. A) Bar chart shows the decrease in median methylation level from 68% to 36% across all 450K probes between control cells and those treated with caged-DAC and UV irradiation. Heatmap shows all normalised beta values plotted in order of chromosome location for the two samples. B) Scatterplot of all normalised beta values for caged-DAC (+UV) against control, showing a distinct pattern of hypomethylation in the caged-DAC (+UV) sample.

5.3.7 Discussion

The potential for achieving targeted demethylation using an *in vitro* culture system was evaluated using two alternative caged nucleoside analogues containing a photoprotective group that may be cleaved through UV irradiation at a particular wavelength, thereby releasing the active molecule. The cytotoxicity of the compounds was assessed alongside their effect on DNA methylation.

The caged-aza molecule was found to be much more toxic to cells in culture than the caged-DAC molecule at equivalent concentrations after irradiation, resulting in almost complete cell death at a final concentration of 1.5µM. Caged-DAC was well tolerated by the cells and produced no noticeable reduction in cell growth rate at 1.5µM when compared to either the control condition or the equivalent non-irradiated treatment. The same observation is made when comparing unmodified 5-aza versus DAC treatment: cells tolerated concentrations of up to 4.5µM DAC (Figure 5-22A), which caused an approximately 40% reduction in final cell confluence, whilst just 1.5µM 5-aza was sufficient to cause a 60% reduction in final cell confluence (Figure 5-16A). As this observation was true of both unmodified and caged versions of these two nucleoside analogues, it is unlikely to be a result of slower uncaging kinetics of the caged-DAC molecule or weaker absorption at 365nm. Furthermore, both possibilities were ruled out after further analysis of the absorption spectra and photolysis dynamics of the two compounds when irradiated (carried out by a collaborator). The differing toxicity may be explained by 5-aza's ability to incorporate into newly synthesised RNA as well as DNA. It has been previously reported that up to 80-90% of 5-aza is preferentially incorporated into RNA as it is a ribonucleoside and cytidine analogue (Li et al., 1970), blocking protein synthesis and triggering apoptosis. This may also explain why significant cell death was seen with the caged-aza treatments in combination with UV irradiation but was not associated with a corresponding decrease in DNA demethylation when the isolated DNA was profiled using the 450K microarray.

The caged-DAC molecule showed significant advantages over the caged-aza molecule, including faster uncaging kinetics, stronger absorption at 365nm and reduced cytotoxicity. Critically, caged-DAC treatment in combination with UV

resulted in a significant reduction in global DNA methylation levels with only a minor reduction in cell growth rate. After a 5-day course of treatment, median DNA methylation levels decreased by 47% compared to the control condition (Figure 5-24), whilst the treated cells still achieved confluence levels of approximately 90% (Figure 5-21B). Unfortunately, the caged-DAC molecule appears to be fairly unstable in aqueous solution and prone to spontaneous hydrolysis, releasing active DAC in the absence of irradiation and reducing median methylation levels by 25%. In order for true temporal control to be achieved over the action of such demethylating drugs, alternative phototags or DAC binding sites need to be explored in order to ensure long-term stability of the drug in its inactive form.

Chapter 6 Discussion

6.1 5hmC: a novel epigenetic modification in mammalian DNA

Since the rediscovery of 5hmC as a novel epigenetic modification in mammalian DNA in 2009, there has been intense interest in the potential role of 5hmC in both transcriptional regulation and DNA demethylation dynamics. The TET proteins were identified as mediators of the oxidative conversion of 5mC to 5hmC and were shown to catalyse the formation of two further oxidative derivatives of 5hmC (5fC and 5caC), leading to the hypothesis that 5hmC was an intermediate in an active DNA demethylation pathway (Ito et al., 2011). However, the accumulation of high levels of 5hmC in particular tissues and identification of 5hmC-specific binding proteins suggests that this modification can be relatively stable and may exhibit its own unique gene regulatory functions, much like 5mC (Spruijt et al., 2013; Yildirim et al., 2011).

Initial publications focused on quantifying global 5hmC levels in a variety of mammalian tissues before attempting to map the distribution of this novel modification throughout the genome in order to gain insight into its possible physiological roles. Tissues of the CNS were consistently found to contain the highest global 5hmC content, around five to ten times higher than all other somatic tissues or stem cell lines, where it can represent up to 40% of the modified cytosine content of DNA (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Song et al., 2011). In addition, 5hmC appeared depleted in highly proliferative tissues and cell lines, including blood and human cancers, suggesting an inverse relationship between 5hmC and proliferative capacity. Many of the first genome-wide profiles used affinity-based methods, such as hMeDIP, GLIB and CMS enrichment strategies, to map 5hmC in ESCs at moderation resolution and in a semi-quantitative manner (Ficz et al., 2011; Pastor et al., 2011; Robertson et al., 2011; Stroud et al., 2011; Szulwach et al., 2011a; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011).

As a result of the observation of uniquely high 5hmC levels in the mammalian CNS, focus shifted to 5hmC profiling in tissues of the brain, including mouse cerebellum, hippocampus and human frontal cortex samples (Jin et al., 2011a; Song et al., 2011; Szulwach et al., 2011b). Dot blot analyses of DNA extracted from mouse cerebellum

at several developmental time points revealed a substantial increase of global 5hmC with age, from 0.10% of total nucleotides in the postnatal day 7 (P7) brain to 0.41% in the adult (10-month-old) brain. The P7 mouse brain is composed mainly of proliferative cells undergoing developmental maturation, further strengthening the hypothesis that 5hmC is acquired during neural development and cellular differentiation. 5hmC levels in cultured mESCs and mNSCs were measured at 0.047% and 0.036% of total nucleotides, respectively, just over 50% lower than observed in P7 cerebellum, consistent with the notion that these cells are at a less developmentally mature, pluripotent state (Song et al., 2011). Genome-wide profiling of 5hmC in adult mouse cerebellum using ß-GT labelling, enrichment and sequencing revealed that 5hmC was highly enriched in gene bodies and depleted at the TSS and more distal genomic regions, in contrast to 5mC (as profiled using MeDIP-seq), where around 80% of all reads mapped to repeat sequences and satellite DNA regions. Localised enrichment of 5hmC at proximal upstream and downstream regions surrounding the TSS was also observed (up to approximately 875bp upstream and 200bp downstream). 5hmC levels at intragenic and proximal promoter regions were also positively correlated to gene expression levels in adult cerebellum, suggesting a role for 5hmC in transcriptional regulation.

The first hMeDIP-based genome-wide 5hmC profile in neural tissue investigated the distribution of 5hmC in human frontal cortex a using NimbleGen 720K CpG island (CGI) plus promoter microarray (Jin et al., 2011a). Significant 5hmC enrichment was found in the gene body, whilst a slight depletion was observed at intergenic CGIs and promoters, as previously observed by Song et al. (2011) using an alternative 5hmC enrichment strategy. Comparison of 5hmC to 5mC distribution (mapped using a MIRA assay coupled to microarray hybridisation) showed that 5hmC is much more frequently targeted to promoter regions than 5mC, whilst the latter is enriched in both intergenic and intragenic CGIs. 5hmC was enriched the promoter of genes associated with neural development and patterning processes and showed a positive correlation with gene expression levels when present at low CpG-density (but not intermediate- or high-CpG density) promoters, indicating that 5hmC generation at these regions may promote transcription by relieving the repressive effect of 5mC through a DNA demethylation mechanism.

Further investigation of the genome-wide distribution of 5hmC during neural development was performed through ß-GT labelling, enrichment and sequencing of mouse cerebellum and hippocampus samples at three distinct developmental time points: P7, 6 weeks and 1 year of age (Szulwach et al., 2011b). Overall, 5hmC was enriched within the gene body and gene proximal regions (875bp upstream to 200bp downstream of TSS), but strongly depleted at the TSS and more distal regions. Although hierarchical clustering of the genome-wide profiles showed both cerebellum-specific and hippocampus-specific 5hmC enrichment, all profiles across the three time points in both tissues were highly correlated to each other ($0.81 \leq r \geq 0.96$), suggesting 5hmC is a stable modification whose genome-wide patterns are mostly conserved through development.

Recent evidence for the existence of 5hmC as a stable modification comes from single-base-resolution profiling of 5hmC, 5fC and 5caC within mESCs. A method for single-base-resolution, simultaneous 5fC and 5caC profiling, termed M.SssI-assisted bisulphite sequencing (MAB-seq), has recently been demonstrated in mESCs (Wu et al., 2014). Integration of this dataset with a comparable 5hmC profile generated by TAB-seq (Yu et al., 2012) revealed that 5hmC and 5fC/5caC are predominantly found at different CpG sites. Only 7.5% of all CpGs associated with 5hmC (n=1,571,715) were also associated with 5fC/5caC (n=117,327), suggesting stable 5hmC accumulation occurs at these sites, whilst 81.2% of all 5fC/5caC loci were not associated with the 5hmC modification (n=508,261). It appears that not all 5hmC generation represents an intermediary stage in active DNA demethylation, supporting an additional functional role for 5hmC within the genome.

Due to the intrinsic link between 5mC and 5hmC abundance throughout the genome, profiling both modifications in parallel is key to understanding DNA methylation dynamics during a particular developmental transition. The central aim of this project was to investigate the changes in 5mC and 5hmC distribution across multiple time points during neural differentiation, which had not been previously investigated at the time. It was thought that the substantial gains in 5hmC observed in neural tissue (compared to ESCs and other pluripotent cell types) must have an important functional role during normal development and that precise mapping of these

changes in relation to 5mC would shed light on the possible significance of 5hmC accumulation in the CNS. Previous 5hmC profiles of neural cell types had been made using DNA obtained from heterogeneous tissue samples containing a diverse array of cell types. As epigenetic states vary enormously depending upon cell type, developmental stage, proliferative capacity, disease status and environmental influences, heterogeneous samples can confound downstream analysis and result in misinterpretation of the observed signals. An *in vitro* culture model of neural differentiation was therefore established for the collection of homogeneous cell populations for dual 5mC/5hmC profiling.

6.2 Global 5hmC levels in cultured cells

ESCs have traditionally been cultured in serum- and LIF-containing expansion media, but these cells exhibit both molecular and transcriptional heterogeneity. Heterogeneous expression of the pluripotency transcription factors *Nanog*, *Rex1* and *Stella* are observed under serum- and LIF-containing conditions, leading to variable differentiation potential (Chambers et al., 2007; Hayashi et al., 2008; Toyooka et al., 2008). The existence of interconvertible subpopulations within ESC cultures with fluctuating transcriptional states, once thought to be an inherent characteristic of pluripotent stem cells, is now seen as a destabilisation of the core pluripotency network and consequence of the cell culture environment (Graf and Stadtfeld, 2008; Wray et al., 2010).

More recently, alternative ESC culture conditions have been defined that use three small molecule inhibitors (known as "3i") to block GSK-3-mediated, FGF receptor and ERK signalling, dispensing of the need for exogenous stimuli (growth factors, cytokines or serum constituents) to maintain the pluripotent state through induction of inhibitor-of-differentiation protein expression (Ying et al., 2008). The use of 3i-containing media maintains ESCs in a basal, ground state of self-renewal and pluripotency without the need for serum or feeder cells and exhibit homogeneously high expression levels of both *Nanog* and *Rex1* (Wray et al., 2010). Recent characterisation of the global transcriptome of ESCs grown in either serum-containing media or "2i" conditions showed that the two populations have distinct gene expression profiles (Marks et al., 2012). 2i represents an alternative media

condition using two inhibitors in combination with LIF, resulting in comparable ESC cultures to those grown in 3i (Ying et al., 2008). Over 3,000 genes showed differing transcription levels (>2-fold difference) between the two culture conditions, including 461 genes expressed only in serum and 160 expressed only in 2i. Genes upregulated in serum are enriched for those involved in developmental processes and lineage specification, such as *Pax6*, *Brachyury* and *Runx1*. In addition, analysis of several histone modifications by ChIP-seq under both media conditions showed differing distributions of the repressive histone modification H3K27me3 at the promoters of silent genes and *Hox* clusters, whilst H3K4me3, H3K9me3 and H3K36me3 profiles were very similar across the genome. In particular, substantial reduction of H3K27me3 at the promoters of silent genes in the 2i condition meant that only around 1,000 genes were classified as bivalent in this condition compared to around 3,000 bivalent genes in serum, where bivalency is defined as colocalisation of H3K27me3 and H3K4me3 at the promoter. Taken together, these results suggest that ESCs grown under 2i/3i or serum conditions represent cells at developmentally distinct time points; the use of serum-free culture conditions allows the generation of homogeneous cultures which are more suitable for epigenetic or expression analysis. Cells grown in 3i media were found to show a more homogeneous morphology, uniformly high expression of the pluripotency marker *Oct4 and* reduced spontaneous differentiation compared to those grown in LIF- and serum- containing media.

NSCs have traditionally been expanded in culture within clonal structures termed neurospheres. However, neurospheres exhibit significant cellular heterogeneity and include various differentiated neural cell types in addition to progenitor cells, making them unsuitable for epigenetic or expression analysis (Bez et al., 2003; Suslov et al., 2002). For the purposes of this project, NSCs were instead cultured as adherent monolayers in serum-free media in the presence of EGF and FGF-2, which has been shown to result in homogeneous NSC cultures, as defined by various morphological and molecular markers, all with tripotent differentiation capacity (Conti et al., 2005; Glaser et al., 2007). Both primary and embryonic mNSCs showed uniformly high *nestin* expression, a bipolar morphology and grew in a characteristic lattice-like formation within the culture dish.

Finally, *in vitro*-derived astrocytes were chosen for analysis of a terminally differentiated neural cell type due to the ease of producing large numbers of homogenous, GFAP-positive astrocyte cultures from NSCs via serum exposure; a differentiation for 48 hours is sufficient to produce functionally mature astrocytes (Brunet et al., 2004). Two studies comparing the transcriptomes of *in vitro*-derived astrocytes (using either 10% serum, BMP2, CNTF or PACAP exposure over seven days) to both *in vivo*-derived astrocytes and a variety of other primary neural cells revealed that all astrocytes shared a common expression profile distinct to all other samples (Bachoo et al., 2004; Cahoy et al., 2008), supporting the conclusion that such *in vitro* protocols can produce truly differentiated functional astrocytes comparable to their *in vivo* counterparts. However, hierarchical clustering of the various gene expression profiles showed that that *in vitro*-derived astrocytes consistently formed a distinct subgroup within the astrocytic population. Although some differences between the transcriptomes of *in vivo* and *in vitro* astrocytes were observed, they all showed clearly distinct expression profiles from neurons, oligodendrocytes, neural stem cells and CNS tissues. It has been suggested that *in vitro*-derived astrocytes may not represent fully mature astrocytes and instead may be more similar to immature *in vivo* astrocytes, although the precise reasons behind these differences have not been fully elucidated.

The discovery that *in vitro* differentiated astrocytes (after a 14-day exposure to serum) had similar levels of global 5hmC to mESCs, at around 0.06-0.08% of total nucleotides, rather than the five- to ten-fold higher levels found in primary CNS tissues, suggests that i*n vitro* differentiation protocols may not accurately recapitulate *in vivo* developmental processes (see Chapter 4, Figure 4-14). It is possible that even if the cells present the morphological characteristics of terminally differentiated cells, they may still possess the epigenetic profile of immature astrocytes or glial progenitors; two weeks of differentiation may not sufficient for the phenotypic changes to be reflected in the methylome and/or hydroxymethylome. It is also possible that *in vivo* primary astrocytes have constitutively low levels of 5hmC, as observed in the *in vitro* culture model. The high levels of global 5hmC found in CNS tissue may not be a shared characteristic of all neural cell types, but instead may be a feature unique to neurons. The global 5hmC content of isolated primary neurons is

consistently found to be high, in the range normally associated with the mammalian brain (Hahn et al., 2013; Kriaucionis and Heintz, 2009), but global 5hmC measurements have not yet been made in purified astrocyte or oligodendrocyte populations for comparison. Interestingly, *in vitro* terminally differentiated dopamine neurons have been found to contain a similarly low 5hmC level to that of neural progenitors (and lower than mESCs), suggesting that reduced 5hmC is primarily a consequence of cell culture conditions, rather than an inherent characteristic of the cell type in question (Kim et al., 2014). If this is the case, 5hmC accumulation within primary astrocytes cannot be ruled out.

Both embryonic mNSCs and primary mNSCs (cultured *in vitro*) showed a reduction in global 5hmC levels of approximately 50% in comparison to both mESCs and astrocytes, with 5hmC comprising 0.02% to 0.04% of total nucleotides. This observation is in line with several other publications that have also found a significant decrease in 5hmC levels upon ESC differentiation to neural progenitors (Kim et al., 2014; Song et al., 2011; Tan et al., 2013). However, liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of nestin-positive, primary mNSCs isolated directly from the mouse brain (without *in vitro* culture) revealed that 5hmC represented approximately 0.15% of all cytosines, over ten-fold higher than the levels of cultured mNSCs (Hahn et al., 2013).

Reduced 5hmC upon adaptation to cell culture has previously been noted for non-neural tissues, suggesting that *in vitro* culture systems do not accurately reflect the true hydroxymethylome of primary cell types (Nestor et al., 2012). 5hmC reduction appeared to be a result of *Tet1-3* downregulation after adaption to cell culture and 5hmC continued to decrease with each subsequent passage, without any concomitant change to 5mC levels. Reduced 5hmC levels are also observed in cancer cell lines compared to the corresponding normal tissue, whilst immunohistochemical analysis of various tissues show reduced 5hmC staining in stem cell compartments compared to surrounding differentiated tissue (Haffner et al., 2011; Jin et al., 2011b; Kraus et al., 2012; Lian et al., 2012; Orr et al., 2012). Low 5hmC may be a general feature of highly proliferative cells, perhaps due to the existence replication-dependent, passive

DNA demethylation process that allows complete 5hmC turnover to unmodified cytosine.

6.3 Genome-wide 5hmC profiling during neural development

The genome-wide distribution of 5mC and 5hmC was profiled in mESCs, mNSCs and *in vitro* differentiated astrocytes using MeDIP-seq and hMeDIP-seq. The vast majority of DMRs and hDMRs (97.7% and 99.3% of all DMRs and hDMRs, respectively) were identified between mESCs and mNSCs, suggesting that the most significant changes in DNA methylation occur early on in development, during the initial lineage commitment of mESCs to neural precursors. This is consistent with findings concerning hESC differentiation, where the majority of observed DNA methylation changes (both 5mC and 5hmC) were found during the early developmental transition to hNPCs rather than during the terminal differentiation step to neurons (Kim et al., 2014).

However, nearly 20-fold fewer hDMRs were identified overall than DMRs (323 compared to 6,065), which may reflect the reduced abundance of 5hmC in the genome. It is possible that 5hmC undergoes rapid turnover to further oxidative derivatives, such as 5fC or 5caC, or unmodified cytosine, making it difficult to capture sites of active DNA demethylation through 5hmC profiling. The low absolute levels of 5hmC at any given CpG, combined with the fact that 5hmC is preferentially found in regions of low CpG density (Yu et al., 2012), suggests that antibody-based immunoprecipitation is not ideally suited to 5hmC profiling due to bias towards highly-modified and/or CpG-dense genomic regions. The latter was supported by the observation that around 80% of all identified hDMRs fell within repetitive regions of the genome (see Chapter 4, Figure 4.3). Several published hMeDIP-generated 5hmC profiles of mESCs have also identified significant enrichment within repetitive DNA and a strong bias towards CA/CT repeats (Ficz et al., 2011; Matarese et al., 2011; Williams et al., 2011), despite the fact that more recently developed methods of single-base-resolution profiling have found that 5hmC is largely absent from repeat and satellite DNA and is found almost exclusively in the CpG context in both ESCs and brain tissue (Lister et al., 2013; Wen et al., 2014; Yu et al., 2012). It appears that the high enrichment within simple

repeats seen in hMeDIP-based profiles is an artefact of the enrichment method, likely resulting from cross-reactivity of the anti-5hmC antibody against unmodified cytosine in cytosine-rich regions. Interestingly, a hMeDIP-seq profile of primary mNPCs did not observe a bias towards CA repeats in their analysis, but significantly higher global 5hmC (approximately 0.15% of total nucleotides) was reported in this sample in comparison to the cultured cells studied here, suggesting that the hMeDIP-seq protocol may not be sufficiently optimised to efficiently enrich for low-abundance 5hmC but is effective in samples with high global 5hmC levels (Hahn et al., 2013). Genome-wide TAB-seq profiling of both ESCs and prefrontal cortex tissue has highlighted the discrepancy in 5hmC abundance between these samples: just over 2 million 5hmC bases were identified in E14 mESCs, whilst 28.4 million 5hmC bases were identified in human adult prefrontal cortex (Wen et al., 2014; Yu et al., 2012). Analysis of human foetal prefrontal cortex (aged 22 weeks) found similar 5hmC levels to mESCs (approximately 2.6 million 5hmC bases), perhaps reflecting of the highly proliferative state of the foetal brain.

Approximately equal numbers of hypo- and hyper-hDMRs were identified during the mESC to mNSC transition. No particular patterns of enrichment were seen for the hyper-hDMR set, but hypo-hDMRs were significantly enriched at CGIs and CpG shores (see Chapter 4, Figure 4.1). A similar enrichment was observed by Szulwach et al., who profiled 5hmC during mouse brain development (between postnatal day 7, 6-week and 1-year-old mice) using ß-GT labelling, enrichment and high-throughput sequencing (Szulwach et al., 2011b). They found that hDMRs were enriched at non-TSS CGIs, as well as being enriched more generally for both CpG frequency and GC content.

Comparison of the full 5hmC peak set against a randomised control set for each biological cohort showed 5hmC enrichment only within repeat regions of the genome, whilst 5hmC distribution within enhancers, intragenic regions and transcription factor binding sites mirrored the distribution expected by chance alone (Supplementary Figure 18). No 5hmC enrichment was found within the gene body or enhancer elements, unlike many other reported 5hmC profiles (Ficz et al., 2011; Kim et al., 2014; Pastor et al., 2011; Szulwach et al., 2011b; Tan et al., 2013; Xu et al.,

2011; Yu et al., 2012). Furthermore, EpiExplorer analysis of the hDMR coordinates showed depletion at enhancer regions (marked by H3K4me1 occupancy), in disagreement with another study that reported significant fold changes in 5hmC distribution at enhancers and other distal regulatory elements between ESCs and neural progenitors (Kim et al., 2014). Kim et al. found that during the hESC to hNPC transition, the largest fold changes in 5hmC enrichment were found at exons and enhancers (approximately 8-fold), whilst smaller but significant changes were also observed in promoters. Similarly, single-base-resolution analysis using BS- and TAB-seq in both hESCs and mESCs has found that nearly half (46.4%) of all identified 5hmC bases are located within promoter-distal regulatory elements, and the highest absolute 5hmC levels are also found at these sites, most notably at poised enhancers (marked by H3K4me1 only) (Yu et al., 2012). 5hmC enrichment within enhancer regions has also been observed in human adult prefrontal cortex, with poised enhancers displaying the highest genomic 5hmC level, at 32.6% (Wen et al., 2014). These results support a regulatory role for 5hmC in gene transcription, possibly by mediating DNA demethylation events at these functionally important genomic regions or altering the DNA-binding affinity of regulatory proteins.

In contrast to 5hmC, the majority of DMRs called between mESCs and mNSCs fell within intragenic regions and were significantly enriched within exons and depleted in repeat regions (see Chapter 4, Figure 4-1A). EpiExplorer analysis revealed that DMRs identified between mESCs and mNSCs were extensively enriched at gene regulatory sites, including those marked by H3K4me1, H3K4me3, DNaseI hypersensitivity and CTCF binding, in addition to evolutionarily conserved regions (see Chapter 4, Figure 4-2). Hyper-DMRs also showed enrichment at Pol-II binding sites, promoters (-5kb to 1kb) and CGIs, although the latter category was not found to contain statistically significant DMR enrichment during MeDUSA analysis of the MeDIP-seq datasets. Overall, these results suggest that the identified DNA methylation changes have functional relevance during this initial transition towards neural lineage commitment. Likewise, GO analysis of DMR-associated gene sets revealed enrichment for genes associated with developmental processes, such as embryonic morphogenesis and neural tube patterning, as well as genes upregulated in neural stem cells and brain tissue.

Approximately 40-fold fewer DMRs were called during the mNSC to astrocyte transition in comparison to the mESC to mNSC transition, all representing gains in 5mC. These hyper-DMRs did not show significant enrichment for any particular gene feature and further analysis using EpiExplorer supported this conclusion, finding no enrichment in alternative genomic features; the distribution of hyper-DMRs within various genomic features followed the same pattern as expected by chance. However, GO analysis of the hyper-DMR-associated genes did find significant enrichment for genes upregulated in various CNS tissues, such as *Cntnap2*, pointing to a biologically significant role for these DMRs in astrocytic differentiation.

No overlap was found between hypo-DMR coordinates and hyper-hDMRs during the astrocytic differentiation of mESCs, made difficult by the low numbers of hDMRs observed overall in comparison to DMRs (323 versus 6,065), preventing any direct observation of a dynamic relationship between these two modifications. Many other regions of differential 5hmC content are likely to have been overlooked due to an inability to immunoprecipitate DNA fragments with low CpG content. The identification of hDMRs was also constrained by the overrepresentation of repetitive DNA sequences in the hMeDIP DNA libraries, resulting in a high 'background' 5hmC signal that masked small but significant changes in 5hmC distribution. Moreover, regions with low absolute levels of 5hmC, even if present at high density in CGIs, may have been missed due to insufficient sequencing depth. It is now apparent that hMeDIP-seq profiling is not ideally suited to profiling 5hmC, unless extensive optimisations are made to compensate for the low abundance of this modification in comparison to 5mC. Base-resolution 5hmC analysis of mESCs within CGIs, using reduced representation oxBS-seq, found that just 800 CGIs (~6%) contained detectable 5hmC (out of 12,660 covered by sequencing) (Booth et al., 2012). The average 5hmC level across these CGIs was just 3.3% (at a maximum of 18.5%) and was inversely correlated to CpG density. Similarly, genome-wide TAB-seq profiling of mESCs also found an inverse correlation between 5hmC and CpG content, highlighting the difficulty in profiling 5hmC using an enrichment method biased towards CpG-dense regions such as CGIs (Yu et al., 2012).

6.4 Developments in 5hmC profiling

At the start of this project, methods for analysing genome-wide 5hmC were not well established. Popular methods of 5mC detection, including 450K BeadChip analysis and single-base-resolution sequencing assays, were not easily adapted for 5hmC analysis because they rely on BS conversion of DNA, which cannot discriminate between 5hmC and 5mC within DNA molecules (Huang et al., 2010; Nestor et al., 2010). Genome-wide 5hmC profiling was restricted to affinity enrichment-based assays coupled to either sequencing or microarray analysis, but these methods offer only moderate resolution, may show bias or cross-reactivity to cytosine-dense region, as in the case of antibody-based methods, or may require custom-made reagents, as in the case of the ß-GT selective chemical labelling method implemented by Song et al. (2011).

The development of two single-base-resolution 5hmC profiling methods, TAB-seq and oxBS-seq, have dramatically increased our understanding of 5hmC distribution within the genome, although the considerable cost of high-throughput sequencing has limited the wider application of these methods (Booth et al., 2012; Yu et al., 2012). The development of oxBS conversion as a method of distinguishing 5hmC from 5mC opened up the possibility of adapting exiting BS-dependent protocols for 5hmC analysis (Booth et al., 2012). One aim of this project was to adapt the oxBS conversion protocol for use with the Infinium HumanMethylation450 BeadChip, allowing the detection of 5hmC and 5mC in parallel at significantly lower cost than high-throughput sequencing. An optimised “high input” protocol was successfully developed (Stewart et al., 2015) that allowed sensitive and reproducible 5hmC detection in human DNA samples in a comparable manner to 5mC detection using the gold standard BS conversion protocol (EZ DNA Methylation Kit from Zymo Research). Technical replicates of oxBS-450K datasets were highly correlated to each other ($r > 0.99$), whilst statistically significant 5hmC levels were detected at individual CpG sites at levels as low as 4%. There was initial concern that the 450K probes may not adequately cover genomic regions associated with 5hmC because 5mC and 5hmC have been observed as having very distinct distributions within the genome. However, application of the oxBS-450K method to a human brain DNA sample highlighted the suitability of the 450K array 5hmC profiling: over 30% of all

CpG loci represented on the array were called as 5hmC, with an mean level of 15% that reached a maximum of 57% at some sites.

In order to measure 5hmC at single-base-resolution in any genomic region of interest, oxBS conversion was subsequently adapted to BS pyrosequencing (oxBS-pyroseq) (Stewart et al., 2015). Pyrosequencing is frequently used for targeted 5mC analysis and is generally considered the gold standard technique for validation of 5mC profiling methods, including MeDIP-seq and 450K arrays (Dejeux et al., 2009; Harris et al., 2010; Roessler et al., 2012). Comparison of the pyrosequencing readout of oxBS-converted and BS-converted DNA amplicons allowed the 5hmC content of individual CpG sites to be determined and was used to successfully validate oxBS-450K datasets, with 5hmC levels measured in the range of 30-50% at these targeted CpG sites. Measurements differed by less than 10% between oxBS-450K and oxBS-pyroseq (corresponding to a difference of less than 5% in absolute 5hmC abundance), suggesting that both represent reliable methods of detecting base-resolution 5hmC within DNA.

More recently, an alternative method of 5hmC detection using the Infinium 450K has been developed that relies on Tet-assisted bisulphite conversion to discriminate 5hmC from 5mC, referred to as TAB-array (Chopra et al., 2014; Nazor et al., 2014). TAB conversion protects 5hmC whilst resulting in the conversion of both 5mC and unmodified C to uracil, allowing 5hmC detection within a DNA sample using just one 450K array. For both 5mC and 5hmC detection, a BS-converted DNA sample must be assayed in parallel, in a similar manner to the oxBS-450K method. However, this TAB-array relies on the production of highly active recombinant TET1 for DNA oxidation, which is both laborious and technically challenging, whilst the oxidation reagent used for oxBS-450K analysis (potassium perruthenate) is commercially available. Nevertheless, both methods offer genome-wide, single-base resolution profiling of 5hmC within human DNA at substantially lower cost than DNA sequencing. In addition, the bioinformatics analysis required to process 450K data is much less complex than that of sequencing datasets, greatly increasing the accessibility of these 5hmC detection methods to researchers.

6.5 Developments in targeted DNA demethylation

Epigenetic deregulation is frequently linked to disease progression and both aberrant 5mC and 5hmC distributions have been associated with cancer (Feinberg and Vogelstein, 1983; Jin et al., 2011b; Jones and Baylin, 2007). Consequently, there is great need to both understand the cause of this deregulation and develop new therapies targeting the DNA methylation and/or demethylation machinery. Acute myeloid leukaemia (AML), which is characterised by the presence of a TET1-MLL oncofusion protein, is frequently associated with either *TET2* or isocitrate dehydrogenase *(IDH1/2)* mutations and a global hypo-hydroxymethylation phenotype, most likely due to disruption of TET-mediated 5mC oxidation (Figueroa et al., 2010; Ko et al., 2010), although 5hmC reduction may be a secondary event to the loss of its substrate, 5mC. Indeed, both global hyper- and hypomethylation are observed in AML samples harbouring *TET2* mutations, suggesting that many other components of the DNA methylation machinery are disrupted during the development of this malignancy, resulting in complex and varying alterations to the methylome.

Two DNMT inhibitors (5-aza and DAC) have been granted FDA approval for the treatment of various myelodysplastic syndromes (MDS), including AML and primary myelofibrosis (see Chapter 5, section 5.3.1). Both drugs have shown effectiveness at reducing aberrant hypermethylation and restoring normal gene expression in cancer cell lines (Bogani et al., 2008; Hagemann et al., 2011; Hollenbach et al., 2010) and improving the overall survival and quality of life of MDS patients (Kuykendall, 2005; Silverman, 2002; Szmigielska-Kaplon and Robak, 2011). In addition to their DNA demethylating activity, these drugs exhibit extensive cytotoxity due to their incorporation into DNA during replication, which results in DNA damage and inhibition of DNA synthesis at high concentrations (Li et al., 1970; Qin et al., 2009; Stresemann and Lyko, 2008). Both cytosine analogues are also non-specific in action and unstable in aqueous solution, with a half-life of only around 20 minutes in blood plasma, making their use in a clinical setting very challenging (Krečmerová and Otmar, 2012). Patients must be given drug infusions over many hours and serious side effects are common due to the drug's effects on healthy tissue. 5-aza, but not DAC, is also readily incorporated into RNA molecules,

substantially increasing its potential cytotoxity due to its ability to alter mRNA processing (and thus gene expression levels) in both proliferative and non-proliferative cells. The development of novel drugs with increased specificity is will be crucial for improving the efficacy of these epigenetic therapies for cancer treatment.

The final aim of this project was to explore the potential of targeted DNA demethylation therapy using photocaged DNMT inhibitors to selectively target cancerous cells in a spatially and temporally controlled manner. This part was conducted in collaboration with Ha Phuong Nguyen and Stefan Howorka from UCL Chemistry, who synthesised two photocaged DNMT inhibitors through the addition of a photosensitive moiety at the N4 position of the pyrimidine ring of both 5-aza and DAC, forming caged-aza and caged-DAC respectively. The phototag renders these molecules biologically inactive but may be cleaved from the pyrimidine ring by UV irradiation, releasing the active cytidine analogue. It was thought that these DNMT inhibitors could be administered to patients in an inactive form and then activated only at the site of diseased tissue through targeted irradiation. The viability of this treatment option was first explored *in vitro* on an osteosarcoma cancer cell line (Saos-2). Varying concentrations of caged-aza and caged-DAC were administered to Saos-2 cells in order to assess cytotoxicity, stability, uncaging rate and DNA demethylating activity in comparison to traditional 5-aza and DAC treatments.

Both caged-aza and caged-DAC treatments resulted in reduced cell viability in a concentration-dependent manner, suggesting they were prone to hydrolysis in cell culture even in the absence of UV irradiation, resulting in the release of active 5-aza and DAC molecules. However, cell viability was reduced further upon irradiation of the treated cells, despite the fact that irradiation on its own had no visible effect on cell growth rate or survival. This observation implied that UV irradiation (at 365nm) resulted in further uncaging of the compounds in cell culture and only partial hydrolysis had occurred under normal growth conditions. Furthermore, comparisons with 5-aza and DAC treatments at equivalent concentrations over a five-day period revealed that the photocaged molecules were far less toxic to the cells than the

traditional DNMT inhibitors, suggesting that only partial uncaging of the active molecule occurred in the culture medium after UV irradiation.

At equivalent concentrations, caged-aza exhibited much greater cytotoxicity than caged-DAC under irradiation conditions. Treatment with 1.5µM caged-aza followed by UV irradiation over the course of five days resulted in almost complete cell death, whereas 1.5µM caged-DAC in combination with UV was well-tolerated by the cells, resulting in only a minor decrease (approximately 10%) in cell confluence compared to controls. This difference could not be explained by the differential uncaging kinetics of the two molecules during irradiation, as HPLC analysis confirmed that caged-DAC actually underwent much faster uncaging to its active form, with over 92% of the compound uncaged after one hour, in compared to just 16% of caged-aza. The higher cytotoxicity of caged-aza must have arisen from the generation of other hydrolytic breakdown products during treatment or the effects of RNA incorporation, rather than simply the release of larger quantities of active 5-aza.

Overall, the highest caged-DAC dose tolerated by the cells was at a 1.5µM final concentration within the culture medium, resulting in a minimal reduction in the overall growth rate. As the DNA demethylation activity of DNMT inhibitors is dependent on DNA replication, the proliferative capacity of the cells needs to be maintained in order to assess the effects of the treatment on DNA methylation. The Infinium 450K array was used for genome-wide 5mC profiling of the Soas-2 cell line before and after 1.5µM caged-DAC treatments, either with or without UV irradiation. A substantial reduction in global 5mC levels was observed after 1.5µM caged-DAC in combination with UV, from a median abundance of 68% 5mC in control cells to 36% 5mC after treatment. The demethylation effects of caged-DAC treatment were not constricted to particular genomic sites but acted genome-wide on the methylome of the Saos-2 cells, resulting in a global hypomethylation phenotype. However, a smaller but nevertheless significant reduction in global 5mC levels (from 68% to 51%) was observed in the non-irradiated, caged-DAC-treated sample, confirming that spontaneous hydrolysis of caged-DAC to its active form occurs within the culture media. Nevertheless, caged-DAC remains a promising new treatment option for epigenetic therapy that has the potential to offer increase

specificity and minimal toxicity compared to the currently available DNMT inhibitors.

Further work is needed in order to achieve true temporal control over the activity of caged-DAC. Firstly, the stability of the caged azanucleoside will need to be increased to ensure the molecules remains inactive in aqueous solution. Secondly, it is currently unknown whether uncaging occurs predominantly inside or outside of the treated cells. Whether or not the cells are capable to taking up the photocaged molecules prior to irradiation will be crucial in determining if the targeted *in vivo* irradiation of solid tumours is a viable treatment option. If strict spatial and temporal control over the site of caged-DAC activation can be demonstrated, the cytotoxic effects of caged-DAC at high doses could also be exploited for use as a chemotherapy agent, selectively targeting malignant tumours through the precise application of UV irradiation.

Intriguingly, a recent publication has reported that 5hmC levels are also affected by DAC treatment in an AML cell line (Chowdhury et al., 2015). Despite a reduction in global 5mC, Chowdhury et al. observe an increase in 5hmC levels (along with 5fC and 5caC) after DAC treatment alongside an increase in the TET1-bound fraction of DNA. As a global reduction in 5hmC levels is frequently observed in cancer, this finding offers a new mechanism by which DAC may be mediating its therapeutic effect and highlights the importance of understanding the dynamic interrelationships between 5mC, 5hmC and their further oxidative derivatives when considering the role of DNA methylation in both health and disease. In conclusion, the work presented in this thesis contributes towards a better understanding of the complex methylome dynamics during neural differentiation, suggests limitations to the use of *in vitro* cell models for studying 5hmC and provides a novel method for the genome-wide detection of this modification at single-base resolution.

References

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144–147.

Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., and Bekman, E. (2009). Neural Differentiation of Embryonic Stem Cells In Vitro: A Road Map to Neurogenesis in the Embryo. PLoS One *4*, e6286.

Alvarez-Buylla, A., and Lim, D.A. (2004). For the Long Run: Maintaining Germinal Niches in the Adult Brain. Neuron *41*, 683–686.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. U. S. A. *90*, 11995–11999.

Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Filho, W.J., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. *513*, 532–541.

Baba, Y., Nosho, K., Shima, K., Huttenhower, C., Tanaka, N., Hazra, A., Giovannucci, E.L., Fuchs, C.S., and Ogino, S. (2010). Hypomethylation of the IGF2 DMR in colorectal tumors, detected by bisulfite pyrosequencing, is associated with poor prognosis. Gastroenterology *139*, 1855–1864.

Bachoo, R.M., Kim, R.S., Ligon, K.L., Maher, E.A., Brennan, C., Billings, N., Chan, S., Li, C., Rowitch, D.H., Wong, W.H., et al. (2004). Molecular diversity of astrocytes with implications for neurological disorders. Proc. Natl. Acad. Sci. U. S. A. *101*, 8384–8389.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat. Biotechnol. *27*, 361–368.

Barker, D., Schafer, M., and White, R. (1984). Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. Cell *36*, 131–138.

Beck, S., and Rakyan, V.K. (2008). The methylome: approaches for global DNA methylation profiling. Trends Genet. *24*, 231–237.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L., and Drohat, A.C. (2006). Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. J. Am. Chem. Soc. *128*, 12510–12519.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., et al. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. Nat. Genet. *44*, 40–46.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum. Mol. Genet. *9*, 2395–2402.

Bez, A., Corsini, E., Curti, D., Biggiogera, M., Colombo, A., Nicosia, R.F., Pagano, S.F., and Parati, E.A. (2003). Neurosphere and neurosphere-forming cells: Morphological and ultrastructural characterization. Brain Res. *993*, 18–29.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042–1047.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. Cell *146*, 866–872.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499–1504.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. *3*, 342–347.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.

Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G., and Meissner, A. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat. Biotechnol. *28*, 1106–1114.

Bocker, M.T., Tuorto, F., Raddatz, G., Musch, T., Yang, F.C., Xu, M., Lyko, F., and Breiling, A. (2012). Hydroxylation of 5-methylcytosine by TET2 maintains the active state of the mammalian HOXA cluster. Nat Commun *3*, 818.

Bogani, C., Ponziani, V., Guglielmelli, P., Desterke, C., Rosti, V., Bosi, A., Le Bousse-Kerdilès, M.-C., Barosi, G., and Vannucchi, A.M. (2008). Hypermethylation of CXCR4 promoter in CD34+ cells from patients with primary myelofibrosis. Stem Cells *26*, 1920–1930.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934–937.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539.

Brena, R.M., Huang, T.H.-M., and Plass, C. (2006). Quantitative assessment of DNA methylation: Potential applications for disease diagnosis, classification, and prognosis in clinical settings. J. Mol. Med. (Berl). *84*, 365–377.

Brinkman, A.B., Simmer, F., Ma, K., Kaan, A., Zhu, J., and Stunnenberg, H.G. (2010). Whole-genome DNA methylation profiling using MethylCap-seq. Methods *52*, 232–236.

Britanova, O., Akopov, S., Lukyanov, S., Gruss, P., and Tarabykin, V. (2005). Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. Eur. J. Neurosci. *21*, 658–668.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 Is a Postmitotic Determinant for Upper-Layer Neuron Specification in the Neocortex. Neuron *57*, 378–392.

Brown, A.M., Tekkök, S.B., and Ransom, B.R. (2004). Energy transfer from astrocytes to axons: The role of CNS glycogen. Neurochem. Int. *45*, 529–536.

Brunet, J.F., Grollimund, L., Chatton, J.Y., Lengacher, S., Magistretti, P.J., Villemure, J.G., and Pellerin, L. (2004). Early Acquisition of Typical Metabolic Features upon Differentiation of Mouse Neural Stem Cells into Astrocytes. Glia *46*, 8–17.

Brunet, J.F., Allaman, I., Magistretti, P.J., and Pellerin, L. (2010). Glycogen metabolism as a marker of astrocyte differentiation. J. Cereb. Blood Flow Metab. *30*, 51–55.

Brunner, A.L., Johnson, D.S., Kim, S.W., Valouev, A., Reddy, T.E., Neff, N.F., Anton, E., Medina, C., Nguyen, L., Chiao, E., et al. (2009). Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res. *19*, 1044–1056.

Brüstle, O., Jones, K.N., Learish, R.D., Karram, K., Choudhary, K., Wiestler, O.D., Duncan, I.D., and McKay, R.D. (1999). Embryonic stem cell-derived glial precursors: a source of myelinating transplants. Science *285*, 754–756.

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R.D., and Horsthemke, B. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nat. Genet. *9*, 395–400.

Butcher, L.M., and Beck, S. (2010). AutoMeDIP-seq: A high-throughput, whole genome, DNA methylation assay. Methods *52*, 223–231.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. *28*, 264–278.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chen, W.G., Chang, Q., Lin, Y., Meissner, A., West, A.E., Griffith, E.C., Jaenisch, R., and Greenberg, M.E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science *302*, 885–889.

Chen, Z.X., Mann, J.R., Hsieh, C.L., Riggs, A.D., and Chédin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J. Cell. Biochem. *95*, 902–917.

Chong, S., and Whitelaw, E. (2004). Epigenetic germline inheritance. Curr. Opin. Genet. Dev. *14*, 692–696.

Chopra, P., Papale, L.A., White, A.T., Hatch, A., Brown, R.M., Garthwaite, M.A., Roseboom, P.H., Golos, T.G., Warren, S.T., and Alisch, R.S. (2014). Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. BMC Genomics *15*, 131.

Choufani, S., Shapiro, J.S., Susiarjo, M., Butcher, D.T., Grafodatskaya, D., Lou, Y., Ferreira, J.C., Pinto, D., Scherer, S.W., Shaffer, L.G., et al. (2011). A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. Genome Res. *21*, 465–476.

Chowdhury, B., McGovern, A., Cui, Y., Choudhury, S.R., Cho, I.-H., Cooper, B., Chevassut, T., Lossie, A.C., and Irudayaraj, J. (2015). The hypomethylating agent Decitabine causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells. Sci. Rep. *5*, 9281.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996–2000.

Cihak, A., Weiss, J.W., and Pitot, H.C. (1974). Characterization of Polyribosomes and Maturation of Ribosomal RNA in Hepatoma Cells Treated with 5-Azacytidine. Cancer Res. *34*, 3003–3009.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215–219.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., et al. (2005). Niche-Independent Symmetrical Self-Renewal of a Mammalian Tissue Stem Cell. PLoS Biol *3*, e283.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Cross, S.H., Charlton, J.A., Nan, X., and Bird, A.P. (1994). Purification of CpG islands using a methylated DNA binding column. Nat. Genet. *6*, 236–244.

Davis, T., and Vaisvila, R. (2011). High Sensitivity 5-hydroxymethylcytosine Detection in Balb/C Brain Tissue. J Vis Exp e2661.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771–784.

Dejeux, E., El abdalaoui, H., Gut, I.G., and Tost, J. (2009). Identification and quantification of differentially methylated loci by the pyrosequencing technology. Methods Mol. Biol. *507*, 189–205.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N. Engl. J. Med. *360*, 2289–2301.

Diogo, M.M., Henrique, D., and Cabral, J.M. (2008). Optimization and integration of expansion and neural commitment of mouse embryonic stem cells. Biotechnol.Appl.Biochem. *49*, 105–112.

Doetsch, F., Caille, I., Lim, D.A., Garcia-Verdugo, J.M., and varez-Buylla, A. (1999). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. Cell *97*, 703–716.

Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M., et al. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat. Biotechnol. *26*, 779–785.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P. V, and Laird, P.W. (2000). MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. *28*, E32.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T. V, Davies, R., Down, T.A., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. *38*, 1378–1385.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., Mccune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. *10*, 2709–2721.

Eng, L.F., Ghirnikar, R.S., and Lee, Y.L. (2000). Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). Neurochem. Res. *25*, 1439–1451.

Feber, A., Wilson, G.A., Zhang, L., Presneau, N., Idowu, B., Down, T.A., Rakyan, V.K., Noon, L.A., Lloyd, A.C., Stupka, E., et al. (2011). Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res. *21*, 515–524.

Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. *13*, 97–109.

Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease. Nature *447*, 433–440.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nat Rev Cancer *4*, 143–153.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89–92.

Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. J. Neurosci. Res. *79*, 734–746.

Feng, J., Fouse, S., and Fan, G. (2007). Epigenetic regulation of neural gene expression and neuronal function. Pediatr. Res. *61*, 58–63.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Field, S.F., Beraldi, D., Bachman, M., Stewart, S.K., Beck, S., and Balasubramanian, S. (2015). Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). PLoS One *10*, e0118202.

Figueroa, M.E., bdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y.S., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell *18*, 553–567.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods *7*, 461–U72.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U. S. A. *89*, 1827–1831.

Fu, Y., Rusznák, Z., Herculano-Houzel, S., Watson, C., and Paxinos, G. (2013). Cellular composition characterizing postnatal development and maturation of the mouse brain and spinal cord. Brain Struct. Funct. *218*, 1337–1354.

Futscher, B.W., Oshiro, M.M., Wozniak, R.J., Holtan, N., Hanigan, C.L., Duan, H., and Domann, F.E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat. Genet. *31*, 175–179.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., et al. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature *455*, 351–357.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489–492.

Gautier, F., Bünemann, H., and Grotjahn, L. (1977). Analysis of calf-thymus satellite DNA: evidence for specific methylation of cytosine in C-G sequences. Eur. J. Biochem. *80*, 175–183.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Andreesen, R., Mackensen, A., and Rehli, M. (2006a). Rapid and sensitive detection of CpG-methylation using methyl-binding (MB)-PCR. Nucleic Acids Res. *34*, e82.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. (2006b). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res. *66*, 6118–6128.

Gius, D., Cui, H., Bradbury, C.M., Cook, J., Smart, D.K., Zhao, S., Young, L., Brandenburg, S.A., Hu, Y., Bisht, K.S., et al. (2004). Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. Cancer Cell *6*, 361–371.

Glaser, T., Perez-Bouza, A., Klein, K., and Brüstle, O. (2005). Generation of purified oligodendrocyte progenitors from embryonic stem cells. FASEB J. *19*, 112–114.

Glaser, T., Pollard, S.M., Smith, A., and Brüstle, O. (2007). Tripotential differentiation of adherently expandable neural stem (NS) cells. PLoS One *2*, e298.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010a). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One *5*, e15367.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010b). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One *5*, e15367.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395–398.

Goto, K., Numata, M., Komura, J.I., Ono, T., Bestor, T.H., and Kondo, H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. Differentiation. *56*, 39–44.

Götz, M., and Huttner, W.B. (2005). The cell biology of neurogenesis. Nat. Rev. Mol. Cell Biol. *6*, 777–788.

Graf, T., and Stadtfeld, M. (2008). Heterogeneity of Embryonic and Adult Stem Cells. Cell Stem Cell *3*, 480–483.

Greger, V., Passarge, E., Höpping, W., Messmer, E., and Horsthemke, B. (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum. Genet. *83*, 155–158.

Grunau, C., Hindermann, W., and Rosenthal, A. (2000). Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. Hum. Mol. Genet. *9*, 2651–2663.

Gu, H., Bock, C., Mikkelsen, T.S., Jäger, N., Smith, Z.D., Tomazou, E., Gnirke, A., Lander, E.S., and Meissner, A. (2010). Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat. Methods *7*, 133–136.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat. Genet. *27*, 322–326.

Gyorgy, A.B., Szemes, M., De Juan Romero, C., Tarabykin, V., and Agoston, D. V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur. J. Neurosci. *27*, 865–873.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452.

Haffner, M.C., Chaux, A., Meeker, A.K., Esopi, D., Gerber, J., Pellakuru, L.G., Toubaji, A., Argani, P., Iacobuzio-Donahue, C., Nelson, W.G., et al. (2011). Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget; Vol 2, No 8 August 2011.

Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. (2011). Azacytidine and Decitabine Induce Gene-Specific and Non-Random DNA Demethylation in Human Cancer Cell Lines. PLoS One *6*, e17388.

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P., et al. (2013). Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis. Cell Rep. *3*, 291–300.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*, 15–23.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78–82.

Halachev, K., Bast, H., Albrecht, F., Lengauer, T., and Bock, C. (2012). EpiExplorer: live exploration and global analysis of large epigenomic datasets. Genome Biol. *13*, R96.

Halassa, M.M., Fellin, T., Takano, H., Dong, J.-H., and Haydon, P.G. (2007). Synaptic islands defined by the territory of a single astrocyte. J. Neurosci. *27*, 6473–6477.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y., et al. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat. Biotechnol. *28*, 1097–1105.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. Nat. Genet. *34*, 187–192.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983–1993.

Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391–401.

Haydon, P.G., and Carmignoto, G. (2006). Astrocyte control of synaptic transmission and neurovascular coupling. Physiol. Rev. *86*, 1009–1031.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141–1143.

Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J.P.J., Davidson, N.E., Sidransky, D., and Baylin, S.B. (1995). Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. *55*, 4525–4530.

Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. U. S. A. *93*, 9821–9826.

Herman, J.G., Civin, C.I., Issa, J.P.J., Collector, M.I., Sharkis, S.J., and Baylin, S.B. (1997). Distinct patterns of inactivation of p15(INK4B) and p16(INK4A) characterize the major types of hematological malignancies. Cancer Res. *57*, 837–841.

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J. V., Setien, F., Carmona, F.J., et al. (2012). Distinct DNA methylomes of newborns and centenarians. Proc. Natl. Acad. Sci. *109*, 10522–10527.

Hollenbach, P.W., Nguyen, A.N., Brady, H., Williams, M., Ning, Y., Richard, N., Krushel, L., Aukerman, S.L., Heise, C., and MacBeth, K.J. (2010). A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. PLoS One *5*.

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163–170.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.

Hong, E.E., Okitsu, C.Y., Smith, A.D., and Hsieh, C.-L. (2013). Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. Mol. Cell. Biol. *33*, 2683–2690.

Horsthemke, B., and Buiting, K. (2006). Imprinting defects on human chromosome 15. In Cytogenetic and Genome Research, pp. 292–299.

Hsieh, J., and Gage, F.H. (2004). Epigenetic control of neural stem cell fate. Curr. Opin. Genet. Dev. *14*, 461–469.

Hsu, C.H., Peng, K.L., Kang, M.L., Chen, Y.R., Yang, Y.C., Tsai, C.H., Chu, C.S., Jeng, Y.M., Chen, Y.T., Lin, F.M., et al. (2012). TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. Cell Rep. *2*, 568–579.

Huang, L.H., Farnet, C.M., Ehrlich, K.C., and Ehrlich, M. (1982). Digestion of highly modified bacteriophage DNA by restriction endonucleases. Nucleic Acids Res. *10*, 1579–1591.

Huang, Y., Pastor, W.A., Shen, Y.H., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. PLoS One *5*.

Huang, Y., Pastor, W.A., Zepeda-Martinez, J.A., and Rao, A. (2012). The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. Nat.Protoc. *7*, 1897–1908.

Iadecola, C., and Nedergaard, M. (2007). Glial regulation of the cerebral microvasculature. Nat. Neurosci. *10*, 1369–1376.

Ichihashi, M., Ueda, M., Budiyanto, A., Bito, T., Oka, M., Fukunaga, M., Tsuru, K., and Horikawa, T. (2003). UV-induced skin damage. Toxicology *189*, 21–39.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett. *583*, 1713–1720.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. *6*, 0037–0051.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. *6*, e1001134.

Inoue, A., and Zhang, Y. (2011). Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. Science (80-.). *334*, 194–194.

Irier, H.A., and Jin, P. (2012). Dynamics of DNA Methylation in Aging and Alzheimer's Disease. DNA Cell Biol. *31*, S – 42 – S – 48.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science (80-.). *333*, 1300–1303.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol. Cell. Biol. *24*, 8862–8871.

Jackson-Grusby, L., Laird, P.W., Magge, S.N., Moeller, B.J., and Jaenisch, R. (1997). Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. Proc. Natl. Acad. Sci. U. S. A. *94*, 4681–4685.

Jacobsen, S.E.E. al. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. *12*, R54.

Jacque, C.M., Vinner, C., Kujas, M., Raoul, M., Racadot, J., and Baumann, N.A. (1978). Determination of glial fibrillary acidic protein (GFAP) in human brain tumors. J. Neurol. Sci. *35*, 147–155.

Jaenisch, R. (1997). DNA methylation and imprinting: Why bother? Trends Genet. *13*, 323–329.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248–251.

Jin, S.-G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. *38*, e125.

Jin, S.G., Wu, X., Li, A.X., and Pfeifer, G.P. (2011a). Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. *39*, 5015–5024.

Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., and Pfeifer, G.P. (2011b). 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res. *71*, 7360–7365.

Jones, P.A., and Baylin, S.B. (2007). The Epigenomics of Cancer. Cell *128*, 683–692.

Jorgensen, H.F. (2006). Engineering a high-affinity methyl-CpG-binding protein. Nucleic Acids Res. *34*, e96–e96.

Josse, J., and Kornberg, A. (1962). Glucosylation of deoxyribonucleic acid. III. alpha- and beta-Glucosyl transferases from T4-infected Escherichia coli. J. Biol. Chem. *237*, 1968–1976.

Jüttermann, R., Li, E., and Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. Proc. Natl. Acad. Sci. U. S. A. *91*, 11797–11801.

Kaiser, J. (2010). Epigenetic Drugs Take On Cancer. Science (80-.). *330*, 576–578.

Kaminskas, E., Farrell, A., Abraham, S., Baird, A., Hsieh, L.S., Lee, S.L., Leighton, J.K., Patel, H., Rahman, A., Sridhara, R., et al. (2005). Approval summary: Azacitidine for treatment of myelodysplastic syndrome subtypes. Clin. Cancer Res. *11*, 3604–3608.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112–U14.

Kantarjian, H., Issa, J.-P.J., Rosenfeld, C.S., Bennett, J.M., Albitar, M., DiPersio, J., Klimek, V., Slack, J., de Castro, C., Ravandi, F., et al. (2006). Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer *106*, 1794–1803.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., et al. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat. Genet. *38*, 149–153.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nat. Struct. Mol. Biol. *19*, 1037–1043.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C., et al. (2006). Comparative isoschizomer profiling of cytosine methylation: The HELP assay. Genome Res. *16*, 1046–1055.

Kiefer, J.C. (2007). Epigenetics in development. Dev. Dyn. *236*, 1144–1156.

Kim, J.-H., Auerbach, J.M., Rodríguez-Gómez, J.A., Velasco, I., Gavin, D., Lumelsky, N., Lee, S.-H., Nguyen, J., Sánchez-Pernaute, R., Bankiewicz, K., et al. (2002). Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. Nature *418*, 50–56.

Kim, M., Park, Y.-K., Kang, T.-W., Lee, S.-H., Rhee, Y.-H., Park, J.-L., Kim, H.-J., Lee, D., Lee, D., Kim, S.-Y., et al. (2014). Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. Hum. Mol. Genet. *23*, 657–667.

Kinney, S.M., Chin, H.G., Vaisvila, R., Bitinaite, J., Zheng, Y., Esteve, P.O., Feng, S., Stroud, H., Jacobsen, S.E., and Pradhan, S. (2011). Tissue specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genome. J. Biol. Chem.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired

hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Koehler, R.C., Roman, R.J., and Harder, D.R. (2009). Astrocytes and the regulation of cerebral blood flow. Trends Neurosci. *32*, 160–169.

Kraus, T.F.J., Globisch, D., Wagner, M., Eigenbrod, S., Widmann, D., Munzel, M., Muller, M., Pfaffeneder, T., Hackner, B., Feiden, W., et al. (2012). Low values of 5-hydroxymethylcytosine (5hmC), the 'sixth base,' are associated with anaplasia in human brain tumors. Int. J. Cancer *131*, 1577–1590.

Krečmerová, M., and Otmar, M. (2012). 5-azacytosine compounds in medicinal chemistry: current stage and future perspectives. Future Med. Chem. *4*, 991–1005.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. *32*, 149–184.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H., and Koike, K. (2012). Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. Cancer Sci. *103*, 670–676.

Kumar, V., and Biswas, D.K. (1988). Dynamic state of site-specific DNA methylation concurrent to altered prolactin and growth hormone gene expression in the pituitary gland of pregnant and lactating rats. J. Biol. Chem. *263*, 12645–12652.

Kuykendall, J.R. (2005). 5-Azacytidine and decitabine monotherapies of myelodysplastic disorders. Ann. Pharmacother. *39*, 1700–1709.

Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. Nat. Rev. Genet. *11*, 191–203.

Langemeijer, S.M.C., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A.P., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat. Genet. *41*, 838–842.

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195–3205.

Leone, D.P., Heavner, W.E., Ferenczi, E. a, Dobreva, G., Huguenard, J.R., Grosschedl, R., and McConnell, S.K. (2015). Satb2 Regulates the Differentiation of Both Callosal and Subcerebral Projection Neurons in the Developing Cerebral Cortex. Cereb. Cortex *25*, 3406–3419.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, W., and Liu, M. (2011). Distribution of 5-hydroxymethylcytosine in different human tissues. J. Nucleic Acids *2011*, 870726.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, H., Liu, H., Corrales, C.E., Risner, J.R., Forrester, J., Holt, J.R., Heller, S., and Edge, A.S.B. (2009a). Differentiation of neurons from neural precursors generated in floating spheres from embryonic stem cells. BMC Neurosci. *10*, 122.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009b). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, L.H., Olin, E.J., Buskirk, H.H., and Reineke, L.M. (1970). Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. Cancer Res. *30*, 2760–2769.

Li, N., Ye, M., Li, Y., Yan, Z., Butcher, L.M., Sun, J., Han, X., Chen, Q., Zhang, X., and Wang, J. (2010a). Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods *52*, 203–212.

Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., et al. (2010b). The DNA Methylome of Human Peripheral Blood Mononuclear Cells. PLoS Biol *8*, e1000533.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., et al. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. Cell *150*, 1135–1146.

Liang, P., Song, F., Ghosh, S., Morien, E., Qin, M., Mahmood, S., Fujiwara, K., Igarashi, J., Nagase, H., and Held, W.A. (2011). Genome-wide survey reveals dynamic widespread tissue-specific changes in DNA methylation during development. BMC Genomics *12*, 231.

Lin, K.T., Momparler, R.L., and Rivard, G.E. (1981). High-performance liquid chromatographic analysis of chemical stability of 5-aza-2'-deoxycytidine. J. Pharm. Sci. *70*, 1228–1232.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Liu, C., Liu, L., Chen, X., Shen, J., Shan, J., Xu, Y., Yang, Z., Wu, L., Xia, F., Bie, P., et al. (2013a). Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. PLoS.One. *8*, e62828.

Liu, S., Qu, Y., Stewart, T.J., Howard, M.J., Chakrabortty, S., Holekamp, T.F., and McDonald, J.W. (2000). Embryonic stem cells differentiate into oligodendrocytes and myelinate in culture and after spinal cord transplantation. Proc. Natl. Acad. Sci. U. S. A. *97*, 6126–6131.

Liu, X., Guangsen, Z., Yi, Y., Xiao, L., Pei, M., Liu, S., Luo, Y., Zhong, H., Xu, Y., Zheng, W., et al. (2013b). Decreased 5-hmC levels are associated with TET2 mutation and unfavorable overall survival in myelodysplastic syndromes. Leuk. Lymphoma 1–18.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., et al. (2013c). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. *31*, 142–147.

Lois, C., and Alvarez-Buylla, A. (1993). Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia. Proc. Natl. Acad. Sci. U. S. A. *90*, 2074–2077.

Loo, D.T., Althoen, M.C., and Cotman, C.W. (1995). Differentiation of serum-free mouse embryo cells into astrocytes is accompanied by induction of glutamine synthetase activity. J. Neurosci. Res. *42*, 184–191.

Lorincz, M.T. (2006). Optimized neuronal differentiation of murine embryonic stem cells: role of cell density. Methods Mol. Biol. *330*, 55–69.

Lu, L.J., and Randerath, K. (1980). Mechanism of 5-azacytidine-induced transfer RNA cytosine-5-methyltransferase deficiency. Cancer Res. *40*, 2701–2705.

Lutz, S.E., Zhao, Y., Gulinello, M., Lee, S.C., Raine, C.S., and Brosnan, C.F. (2009). Deletion of astrocyte connexins 43 and 30 leads to a dysmyelinating phenotype and hippocampal CA1 vacuolation. J. Neurosci. *29*, 7743–7752.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. J. Biol. Chem. *286*, 35334–35338.

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol. *13*, R44.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat. Rev. Genet. *11*, 285–296.

Marks, H., Kalkan, T.., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis-Stewart, A., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590–604.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890–893.

Matarese, F., Carrillo-de Santa Pau, E., and Stunnenberg, H.G. (2011). 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol. Syst. Biol. *7*, 562.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502.

Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol *28*, 1079–1088.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008a). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Meissner, A., Mikkelsen, T.S., Gu, H.C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.L., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008b). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–U91.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell *151*, 1417–1430.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., and Sidransky, D. (1995). 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat. Med. *1*, 686–692.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45–50.

Michels, K.B., Binder, A.M., Dedeurwaerder, S., Epstein, C.B., Greally, J.M., Gut, I., Houseman, E.A., Izzi, B., Kelsey, K.T., Meissner, A., et al. (2013).

Recommendations for the design and analysis of epigenome-wide association studies. Nat. Methods *10*, 949–955.

Mirmohammadsadegh, A., Marini, A., Nambiar, S., Hassan, M., Tannapfel, A., Ruzicka, T., and Hengge, U.R. (2006). Epigenetic silencing of the PTEN gene in melanoma. Cancer Res. *66*, 6546–6552.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393–396.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Mol. Cell *30*, 755–766.

Momparler, R.L., Siegel, S., Avila, F., Lee, T., and Karon, M. (1976). Effect of trna from 5-azacytidine-treated hamster fibrosarcoma cells on protein synthesis in vitro in a cell-free system. Biochem. Pharmacol. *25*, 389–392.

Morris, T.J., Butcher, L.M., Feber, A., Teschendorff, A.E., Chakravarthy, A.R., Wojdacz, T.K., and Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics *30*, 428–430.

Morshead, C.M., Reynolds, B.A., Craig, C.G., McBurney, M.W., Staines, W.A., Morassutti, D., Weiss, S., and Van Der Kooy, D. (1994). Neural stem cells in the adult mammalian forebrain: A relatively quiescent subpopulation of subependymal cells. Neuron *13*, 1071–1082.

Muller, T., Gessi, M., Waha, A., Isselstein, L.J., Luxen, D., Freihoff, D., Freihoff, J., Becker, A., Simon, M., Hammes, J., et al. (2012). Nuclear exclusion of TET1 is associated with loss of 5-hydroxymethylcytosine in IDH1 wild-type gliomas. Am.J.Pathol. *181*, 675–683.

Munzel, M., Globisch, D., Bruckl, T., Wagner, M., Welzmiller, V., Michalakis, S., Muller, M., Biel, M., and Carell, T. (2010). Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. Angew. Chemie-International Ed. *49*, 5375–5377.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat. Chem. Biol. *8*, 751–758.

Nazor, K.L., Boland, M.J., Bibikova, M., Klotzle, B., Yu, M., Glenn-Pratola, V.L., Schell, J.P., Coleman, R.L., Cabral-da-Silva, M.C., Schmidt, U., et al. (2014). Application of a low cost array-based technique - TAB-Array - for quantifying and mapping both 5mC and 5hmC at single base resolution in human pluripotent stem cells. Genomics *104*, 358–367.

Neddermann, P., and Jiricny, J. (1993). The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. J. Biol. Chem. *268*, 21218–21224.

Neddermann, P., and Jiricny, J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. Proc. Natl. Acad. Sci. U. S. A. *91*, 1642–1646.

Nestor, C., Ruzov, A., Meehan, R.R., and Dunican, D.S. (2010). Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques *48*, 317–319.

Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., and Meehan, R.R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. *22*, 467–477.

Ng, C.W., Yildirim, F., Yap, Y.S., Dalin, S., Matthews, B.J., Velez, P.J., Labadorf, A., Housman, D.E., and Fraenkel, E. (2013). Extensive changes in DNA methylation are associated with expression of mutant huntingtin. Proc. Natl. Acad. Sci. U. S. A. *110*, 2354–2359.

Ngô, V., Gourdji, D., and Laverrière, J.N. (1996). Site-specific methylation of the rat prolactin and growth hormone promoters correlates with gene expression. Mol. Cell. Biol. *16*, 3245–3254.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Notari, R.E., and DeYoung, J.L. (1975). Kinetics and mechanisms of degradation of the antileukemic agent 5-azacytidine in aqueous solutions. J. Pharm. Sci. *64*, 1148–1157.

Ogata, K., and Kosaka, T. (2002). Structural and quantitative analysis of astrocytes in the mouse hippocampus. Neuroscience *113*, 221–233.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Orr, B., Haffner, M., Eberhart, C., Hicks, J., Nelson, W., and Yegnasubramanian, S. (2012). Decreased 5hmC is Associated with Neural Progenitor Phenotype in Normal Brain and Shorter Survival in Malignant Glioma. J. Neuropathol. Exp. Neurol. *71*, 547.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr.Biol *10*, 475–478.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394–397.

Payer, B., Chuva de Sousa Lopes, S.M., Barton, S.C., Lee, C., Saitou, M., and Surani, M.A. (2006). Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. Genesis *44*, 75–83.

Pellerin, L., Bouzier-Sore, A.K., Aubert, A., Serres, S., Merle, M., Costalat, R., and Magistretti, P.J. (2007). Activity-dependent regulation of energy metabolism by astrocytes: An update. Glia *55*, 1251–1262.

Pettitt, S.J., Liang, Q., Rairdan, X.Y., Moran, J.L., Prosser, H.M., Beier, D.R., Lloyd, K.C., Bradley, A., and Skarnes, W.C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nat. Methods *6*, 493–495.

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl. *50*, 7008–7012.

Píškala, A., and Šorm, F. (1964). Nucleic acids components and their analogues. LI. Synthesis of 1-glycosyl derivatives of 5-azauracil and 5-azacytosine. Collect. Czechoslov. Chem. Commun. *29*, 2060–2076.

Pollard, S.M., Conti, L., Sun, Y., Goffredo, D., and Smith, A. (2006). Adherent neural stem (NS) cells from fetal and adult forebrain. Cereb.Cortex *16 Suppl 1*, i112–i120.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101–1105.

Powell, E.M., and Geller, H.M. (1999). Dissection of astrocyte-mediated cues in neuronal guidance and process extension. Glia *26*, 73–83.

Psoni, S., Sofocleous, C., Traeger-Synodinos, J., Kitsiou-Tzeli, S., Kanavakis, E., and Fryssira-Kanioura, H. (2012). MECP2 mutations and clinical correlations in Greek children with Rett syndrome and associated neurodevelopmental disorders. Brain Dev. *34*, 487–495.

Qin, T., Jelinek, J., Si, J., Shu, J., and Issa, J.P.J. (2009). Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines. Blood *113*, 659–667.

Rakyan, V.K., Down, T.A., Balding, D.J., and Beck, S. (2011). Epigenome-wide association studies for common human diseases. Nat. Rev. Genet. *12*, 529–541.

Rangam, G., Schmitz, K.-M., Cobb, A.J.A., and Petersen-Mahrt, S.K. (2012). AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. PLoS One *7*, e43279.

Rauch, T., and Pfeifer, G.P. (2005). Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab. Invest. *85*, 1172–1180.

Rauch, T., Li, H., Wu, X., and Pfeifer, G.P. (2006). MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res. *66*, 7939–7947.

Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S.K., Kernstine, K.H., Riggs, A.D., and Pfeifer, G.P. (2007). Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc. Natl. Acad. Sci. U. S. A. *104*, 5527–5532.

Ravanat, J.L., Douki, T., and Cadet, J. (2001). Direct and indirect effects of UV radiation on DNA and its components. J. Photochem. Photobiol. B. *63*, 88–102.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Reynolds, B.A., and Weiss, S. (1992). Generation of Neurons and Astrocytes from Isolated Cells of the Adult Mammalian Central-Nervous-System. Science (80-.). *255*, 1707–1710.

Robertson, A.B., Dahl, J.A., Vågbø, C.B., Tripathi, P., Krokan, H.E., and Klungland, A. (2011). A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. *39*, e55.

Robinson, M.D., Statham, A.L., Speed, T.P., and Clark, S.J. (2010). Protocol matters: which methylome are you actually studying? Epigenomics *2*, 587–598.

Roessler, J., Ammerpohl, O., Gutwein, J., Hasemeier, B., Anwar, S., Kreipe, H., and Lehmann, U. (2012). Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res. Notes *5*, 210.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. *242*, 84–89.

Rougier, N., Bourc'his, D., Molina Gomes, D., Niveleau, A., Plachot, M., Pàldi, A., and Viegas-Péquignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev. *12*, 2108–2113.

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., et al. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res. *21*, 1332–1342.

Sakai, Y., Rawson, C., Lindburg, K., and Barnes, D. (1990). Serum and transforming growth factor beta regulate glial fibrillary acidic protein in serum-free-derived mouse embryo cells. Proc. Natl. Acad. Sci. U. S. A. *87*, 8378–8382.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics *6*, 692–702.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol. *241*, 172–182.

Sato, N., Maehara, N., Su, G.H., and Goggins, M. (2003). Effects of 5-aza-2'-deoxycytidine on matrix metalloproteinase expression and pancreatic cancer cell invasiveness. J Natl Cancer Inst *95*, 327–330.

Sauvageot, C.M., and Stiles, C.D. (2002). Molecular mechanisms controlling cortical gliogenesis. Curr. Opin. Neurobiol. *12*, 244–249.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. *23*, 2625–2638.

Schatz, P., Dietrich, D., and Schuster, M. (2004). Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. Nucleic Acids Res. *32*, e167.

Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeier, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew. Chemie - Int. Ed. *51*, 6516–6520.

Serre, D., Lee, B.H., and Ting, A.H. (2009). MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. *38*, 391–399.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Shiota, K., Kogo, Y., Ohgane, J., Imamura, T., Urano, A., Nishino, K., Tanaka, S., and Hattori, N. (2002). Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. Genes to Cells *7*, 961–969.

Shock, L.S., Thakkar, P. V, Peterson, E.J., Moran, R.G., and Taylor, S.M. (2011). DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. Proc. Natl. Acad. Sci. U. S. A. *108*, 3630–3635.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., and Cedar, H. (1999). DNA methylation represses transcription in vivo. Nat. Genet. *22*, 203–206.

Silverman, L.R. (2002). Randomized Controlled Trial of Azacitidine in Patients With the Myelodysplastic Syndrome: A Study of the Cancer and Leukemia Group B. J. Clin. Oncol. *20*, 2429–2440.

Singer-Sam, J., LeBon, J.M., Tanguay, R.L., and Riggs, A.D. (1990). A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res. *18*, 687.

Singh, V., Sharma, P., and Capalash, N. (2013). DNA methyltransferase-1 inhibitors as epigenetic therapy for cancer. Curr. Cancer Drug Targets *13*, 379–399.

Smith, A.G. (1991). Culture and Differentiation of Embryoinic Stem Cells. J. Tissue Cult. Methods *13*, 89–94.

Sofroniew, M. V., and Vinters, H. V. (2010). Astrocytes: Biology and pathology. Acta Neuropathol. *119*, 7–35.

Song, C.-X., Clark, T.A., Lu, X.-Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2011a). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat. Methods *9*, 75–77.

Song, C.-X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. Nat. Biotechnol. *30*, 1107–1116.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011b). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotech *29*, 68–72.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H., and Held, W.A. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc. Natl. Acad. Sci. U. S. A. *102*, 3336–3341.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146–1159.

Stewart, S.K., Morris, T.J., Guilhamon, P., Bulstrode, H., Bachman, M., Balasubramanian, S., and Beck, S. (2015). oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. Methods *72*, 9–15.

Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int. J. Cancer *123*, 8–13.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. *12*, R54.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J. Biol. Chem. *279*, 27816–27823.

Suslov, O.N., Kukekov, V.G., Ignatova, T.N., and Steindler, D.A. (2002). Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres. Proc. Natl. Acad. Sci. U. S. A. *99*, 14506–14511.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

Suzuki, M., Harashima, A., Okochi, A., Yamamoto, M., Matsuo, Y., Motoda, R., Yoshioka, T., and Orita, K. (2002). Transforming growth factor-beta(1) augments granulocyte-macrophage colony-stimulating factor-induced proliferation of umbilical cord blood CD34(+) cells with an associated tyrosine phosphorylation of STAT5. Exp Hematol *30*, 1132–1138.

Szmigielska-Kaplon, A., and Robak, T. (2011). Hypomethylating Agents in the Treatment of Myelodysplastic Syndromes and Myeloid Leukemia. Curr. Cancer Drug Targets *11*, 837–848.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011a). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat. Neurosci. *14*, 1607–1616.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., et al. (2011b). Integrating 5-Hydroxymethylcytosine into the Epigenomic Landscape of Human Embryonic Stem Cells. PLoS Genet *7*, e1002154.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. *38*, e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., and Butcher, L.M. (2012). Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc. *7*, 617–636.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. U. S. A. *99*, 3740–3745.

Tan, L., Xiong, L., Xu, W., Wu, F., Huang, N., Xu, Y., Kong, L., Zheng, L., Schwartz, L., Shi, Y., et al. (2013). Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. Nucleic Acids Res. *41*, e84.

Temple, S. (2001). The development of neural stem cells. Nature *414*, 112–117.

Thambirajah, A.A., Ng, M.K., Frehlick, L.J., Li, A., Serpa, J.J., Petrotchenko, E. V., Silva-Moreno, B., Missiaen, K.K., Borchers, C.H., Adam Hall, J., et al. (2012). MeCP2 binds to nucleosome free (linker DNA) regions and to H3K9/H3K27 methylated nucleosomes in the brain. Nucleic Acids Res. *40*, 2884–2897.

Tiwari, V.K., Tiwari, V.K., McGarvey, K.M., McGarvey, K.M., Licchesi, J.D.F., Licchesi, J.D.F., Ohm, J.E., Ohm, J.E., Herman, J.G., Herman, J.G., et al. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol. *6*, 2911–2927.

Torres-Padilla, M.E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int. J. Dev. Biol. *50*, 455–461.

Tost, J., and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. Nat. Protoc. *2*, 2265–2275.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909–918.

Tsai, P.-C., and Bell, J.T. (2015). Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int. J. Epidemiol. *44*, 1429–1441.

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res *32*, 4100–4108.

Voutsinos-Porche, B., Bonvento, G., Tanaka, K., Steiner, P., Welker, E., Chatton, J.Y., Magistretti, P.J., and Pellerin, L. (2003). Glial glutamate transporters mediate a functional metabolic crosstalk between neurons and astrocytes in the mouse developing cortex. Neuron *37*, 275–286.

Waalwijk, C., and Flavell, R.A. (1978). MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaII sites. Nucleic Acids Res. *5*, 3231–3236.

Wada, H., Kagoshima, M., Ito, K., Barnes, P.J., and Adcock, I.M. (2005). 5-Azacytidine suppresses RNA polymerase II recruitment to the SLPI gene. Biochem. Biophys. Res. Commun. *331*, 93–99.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. *20*, 116–117.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet *37*, 853–862.

Weiss, S., Dunne, C., Hewson, J., Wohl, C., Wheatley, M., Peterson, A.C., and Reynolds, B.A. (1996). Multipotent CNS stem cells are present in the adult mammalian spinal cord and ventricular neuroaxis. J. Neurosci. *16*, 7599–7609.

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., et al. (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. *15*, R49.

White, H.E., Durston, V.J., Harvey, J.F., and Cross, N.C.P. (2006). Quantitative analysis of SRNPN gene methylation by pyrosequencing as a diagnostic test for Prader-Willi syndrome and Angelman syndrome. Clin. Chem. *52*, 1005–1013.

Wiese, C., Rolletschek, A., Kania, G., Blyszczuk, P., Tarasov, K. V, Tarasova, Y., Wersto, R.P., Boheler, K.R., and Wobus, A.M. (2004). Nestin expression--a property of multi-lineage progenitor cells? Cell. Mol. Life Sci. *61*, 2510–2522.

Williams, R.T., and Wang, Y. (2012). A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2'-deoxycytidines. Biochemistry *51*, 6458–6462.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J. V, Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Wilson, A.S., Power, B.E., and Molloy, P.L. (2007). DNA hypomethylation and human diseases. Biochim. Biophys. Acta *1775*, 138–162.

Wilson, G.A., Dhami, P., Feber, A., Cortázar, D., Suzuki, Y., Schulz, R., Schär, P., and Beck, S. (2012). Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers. Gigascience *1*, 3.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem.Soc.Trans. *38*, 1027–1032.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. *11*, 607–620.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. *25*, 679–684.

Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. Nat. Biotechnol. *32*, 1231–1240.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816–831.

Xinarianos, G., McRonald, F.E., Risk, J.M., Bowers, N.L., Nikolaidis, G., Field, J.K., and Liloglou, T. (2006). Frequent genetic and epigenetic abnormalities contribute to the deregulation of cytoglobin in non-small cell lung cancer. Hum. Mol. Genet. *15*, 2038–2044.

Xiong, Z., and Laird, P.W. (1997). COBRA: A sensitive and quantitative DNA methylation assay. Nucleic Acids Res. *25*, 2532–2534.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187–191.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Mol. Cell *In Press,*.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460–464.

Yang, H., Liu, Y., Bai, F., Zhang, J.-Y., Ma, S.-H., Liu, J., Xu, Z.-D., Zhu, H.-G., Ling, Z.-Q., Ye, D., et al. (2013a). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene *32*, 663–669.

Yang, H., Liu, Y., Bai, F., Zhang, J.Y., Ma, S.H., Liu, J., Xu, Z.D., Zhu, H.G., Ling, Z.Q., Ye, D., et al. (2013b). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene *32*, 663–669.

Yang, X., Lay, F., Han, H., and Jones, P.A. (2010). Targeting DNA methylation for epigenetic therapy. Trends Pharmacol. Sci. *31*, 536–546.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498–1510.

Ying, Q.L., and Smith, A.G. (2003). Defined Conditions for Neural Commitment and Differentiation. Methods Enzymol. *365*, 327–341.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yoder, J.A., and Bestor, T.H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum. Mol. Genet. *7*, 279–284.

Yoshii, A., and Constantine-Paton, M. (2010). Postsynaptic BDNF-TrkB signaling in synapse maturation, plasticity, and disease. Dev. Neurobiol. *70*, 304–322.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012a). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Jin, P., Ren, B., and He, C. (2012b). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat. Protoc. *7*, 2159–2170.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat. Chem. Biol. *8*, 328–330.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell *126*, 1189–1201.

Zhang, Y., Fatima, N., and Dufau, M.L. (2005). Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. Mol. Cell. Biol. *25*, 7929–7939.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Appendices

Appendix 1: Supplementary Tables

Primer pair name	Forward Primer Sequence (5' → 3')	Reverse Primer Sequence (5' → 3')
20CpG (MeDIP QC1)	GGTGAACTTCCGATAGTG	CAGTCATAGATGGTCGGT
15CpG (MeDIP QC1)	TATCACTGTTGATTCTCGC	TATCACTGTTGATTCTCGC
Unmethylated Region (MeDIP QC2)	GGCTAGAACTGACCAGACAG AC	ATCTGTAGCCAATCCTAGAGCA
Methylated Region (MeDIP QC2)	CATGGCCCACAAAGTAATAA AA	AACGACTTACAACGAGCTCAA A
Actin Positive (BS conversion efficiency)	TGGTGATGGAGGAGGTTTAGT AAGT	AACCAATAAAACCTACTCCTCC CTTAA
Actin Negative (BS conversion efficiency)	TGGTGATGGAGGAGGCTCAG CAAGT	AGCCAATGGGACCTGCTCCTCC CTTGA
Grm4 hyper-hDMR (oxBS-pyroseq)	GTTGTTTTGGTAAGTTTAGAA AGTAGTT	*CACTTCCCCTCAACATCTCTA A
Atpase6 hyper-hDMR (oxBS-pyroseq)	AGGTAGATGGTATATTGGTAA TTATGA	*ACCTTTATTCTAATCACAATT CTATCTC
Colgalt2 hyper-hDMR (oxBS-pyroseq)	ATGAAAAGTGATGAGGATTG AATAAG	*AATTACCTCACTATAATATTC ATCTACATC
Jarid2 hyper-DMR (oxBS-pyroseq)	GTTAAAAGGGATTTTAGAGGA GGATT	*CCCTAAACTTTACCTTCCTTA ACC
	oxBS-pyroseq: sequencing primers (5' → 3')	
Grm4 hyper-hDMR	AAGATGTAGGGGTTTT	
Colgalt2 hyper-hDMR	TTATAGATAAAAAAGTATTAATGTT	
Atpase6 hyper-hDMR	ATGAATATTATTATAATTTAATGAG	
Jarid2 hyper-DMR	GGATTTTAGAGGAGGATTT	

Supplementary Table 1: Primer pair sequences. Biotinylated primers are indicated by (*).

MeDIP-seq	ESC1.bed	ESC2.bed	ESC3.bed	NSC1.bed	NSC2.bed	NSC3.bed	AST7.bed	AST8.bed	AST9.bed
ESC1.bed	1.000	0.991	0.987						
ESC2.bed	0.991	1.000	0.996						
ESC3.bed	0.987	0.996	1.000						
NSC1.bed				1.000	0.997	0.995			
NSC2.bed				0.997	1.000	0.995			
NSC3.bed				0.995	0.995	1.000			
AST7.bed							1.000	0.909	0.968
AST8.bed							0.909	1.000	0.868
AST9.bed							0.968	0.868	1.000

hMeDIP-seq	hESC1.bed	hESC2.bed	hESC3.bed	hNSC1.bed	hNSC2.bed	hNSC3.bed	hAST7.bed	hAST8.bed	hAST9.bed
hESC1.bed	1.000	0.994	0.995						
hESC2.bed	0.994	1.000	0.995						
hESC3.bed	0.995	0.995	1.000						
hNSC1.bed				1.000	0.995	0.986			
hNSC2.bed				0.995	1.000	0.989			
hNSC3.bed				0.986	0.989	1.000			
hAST7.bed							1.000	0.959	0.955
hAST8.bed							0.959	1.000	0.980
hAST9.bed							0.955	0.980	1.000

Supplementary Table 2: Correlation matrices of MeDIP-seq profiles (above) and hMeDIP-seq profiles (below).

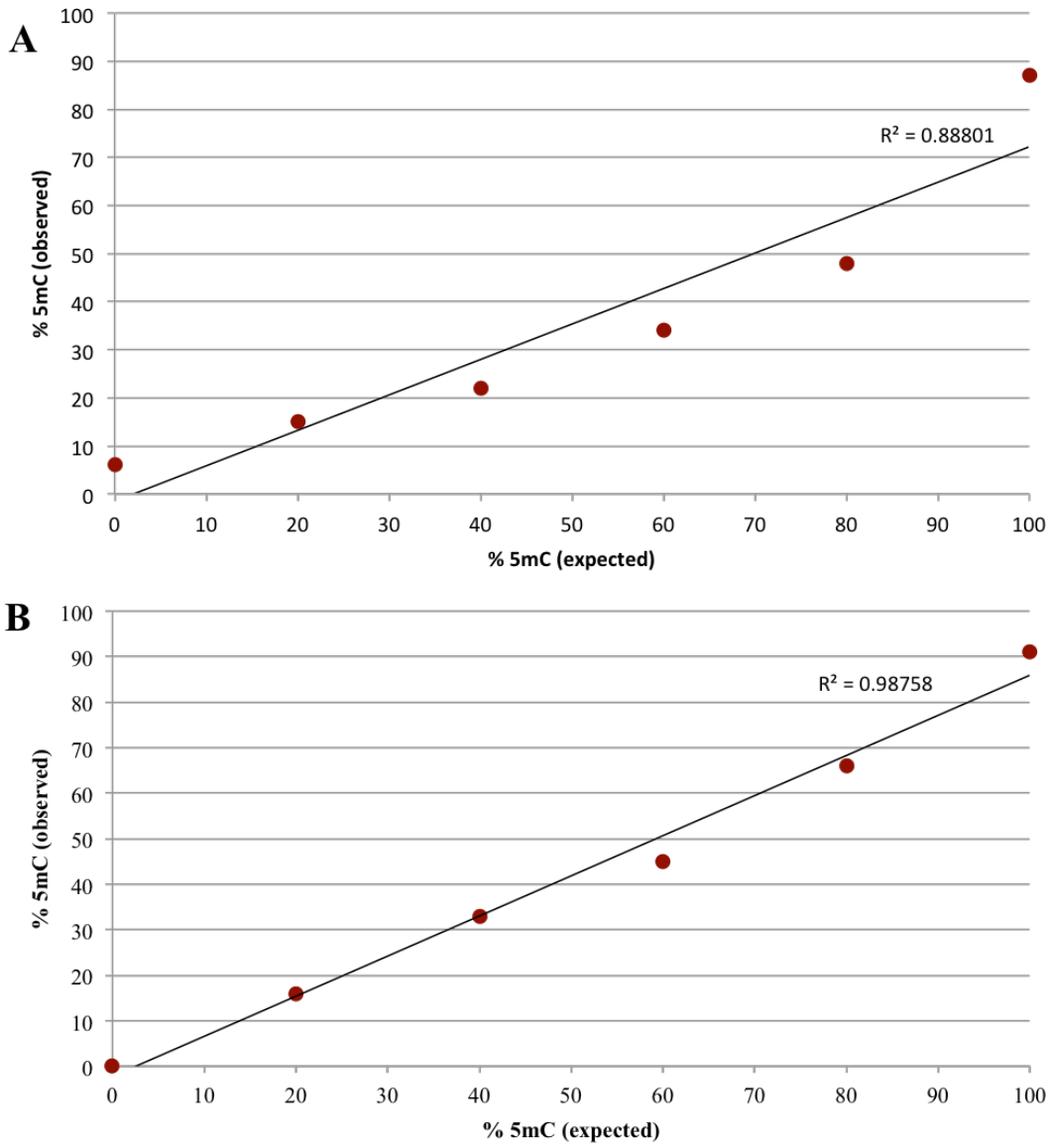
DMR Location	Adjusted P-value	CpG Count	Associated Gene Symbol	Gene Description
chr7: 5030501-5031000	7.01×10^{-29}	48	**Zfp865**	**Zinc finger protein 865**: DNA binding protein involved in regulation of transcription.
chr7:16811001-16811500	1.46×10^{-28}	63	**Fkrp**	**Fukutin related protein:** protein processing within Golgi apparatus.
chr1:56969501-56970000	6.83×10^{-27}	36	**Satb2**	**Special AT-rich sequence binding protein 2:** involved in chromatin remodelling, regulation of transcription and neuron migration.
chr11:69399501-69400000	9.40×10^{-27}	38	**Kdm6b** (also known as Jmjd3)	**KDM1 lysine (K)-specific demethylase 6B:** epigenetic regulator required for normal development and neuronal survival.
chr3:84190501-84191000	3.04×10^{-26}	37	**Trim2** (also known as Narf)	**Tripartite motif-containing 2:** an E3 ubiquitin ligase; regulates neuronal polarization.
chr1:166309001-166309500	5.67×10^{-26}	67	**Ildr2** (also known as Dbsm1)	**Immunoglobulin-like domain containing receptor 2:** involved in cell differentiation and pancreas development.
chr4:136290501-136291000	1.08×10^{-25}	49	**Zfp46** (also known as Znf436)	**Zinc finger protein 46:** DNA binding protein and transcriptional regulator.
chr2:18031501-18032000	1.84×10^{-25}	34	**Gm17762**	**Predicted gene 17762:** non-coding RNA gene.
chr14:54417501-54418000	3.53×10^{-25}	24	**Slc7a7** (also known as my+lat1)	**Solute carrier family 7, member7:** an amino acid transporter; mutations in human ortholog are associated with glioma risk.
chr8:70594501-70595000	4.33×10^{-25}	51	**Isyna1**	**Myo-inositol 1-phosphate synthase A1:** enzyme involved in and myo-inositol (MI) biosynthesis; altered MI levels are associated with multiple brain disorders.

Supplementary Table 3: Top 10 most statistically significant hyper-DMRs identified between mNSCs and mESCs. Associated genes refer to the nearest gene falling within 100kb upstream and 50kb downstream of the DMR. DMRs falling within repeat regions, with a CpG count of zero, were excluded from the list. Gene annotations are taken from the NCBI Gene database and the MGI database, unless otherwise specified.

DMR Location	Adjusted P-value	CpG Count	Associated Gene Symbol	Gene Description
chr8:124765501 - 124766000	1.07×10^{-14}	15	**Fam89a**	**Family with sequence similarity 89, member A:** protein coding gene; function unknown.
chr7:49120001 - 49120500	2.00×10^{-13}	18	**Nav2** (also known as POMFIL2, HELAD1)	**Neuron navigator 2:** encodes a protein involved in cellular growth and migration; associated with nervous system development.
chr8:55346501 - 55347000	1.20×10^{-12}	14	N/A	N/A
chr8:10978501 - 10979000	2.98×10^{-11}	22	**9530052E02Rik**	Long non-coding RNA
chr17:36031501 - 36032000	6.45×10^{-10}	24	**H2-T23** (also known as Qa1)	**Histocompatibility 2, T region locus 23:** gene within the Qa region of the H2 complex; homologous to MHC class I genes.
chr7:48897501 - 48898000	2.12×10^{-09}	4	**E2f8**	**E2F transcription factor 8:** Transcriptional repressor of the E2F family; regulates cell cycle progression; critical for embryonic development.
chr3:119498501 - 119499000	3.02×10^{-09}	16	N/A	N/A
chr5:139051501 - 139052000	3.47×10^{-09}	18	**Prkar1b**	**Protein kinase, cAMP dependent regulatory, type I beta:** Involved in cAMP signalling; required for normal metabolism, ion transport and gene transcription.
chr13:120025001 - 120025500	4.08×10^{-09}	25	**Gm21370**	**Predicted gene 21370**
chr15:78833001 - 78833500	4.08×10^{-09}	12	**Cdc42ep1** (also known as Borg5)	**CDC42 effector protein (Rho GTPase binding) 1:** regulates actin polymerisation and cytoskeleton reorganisation.

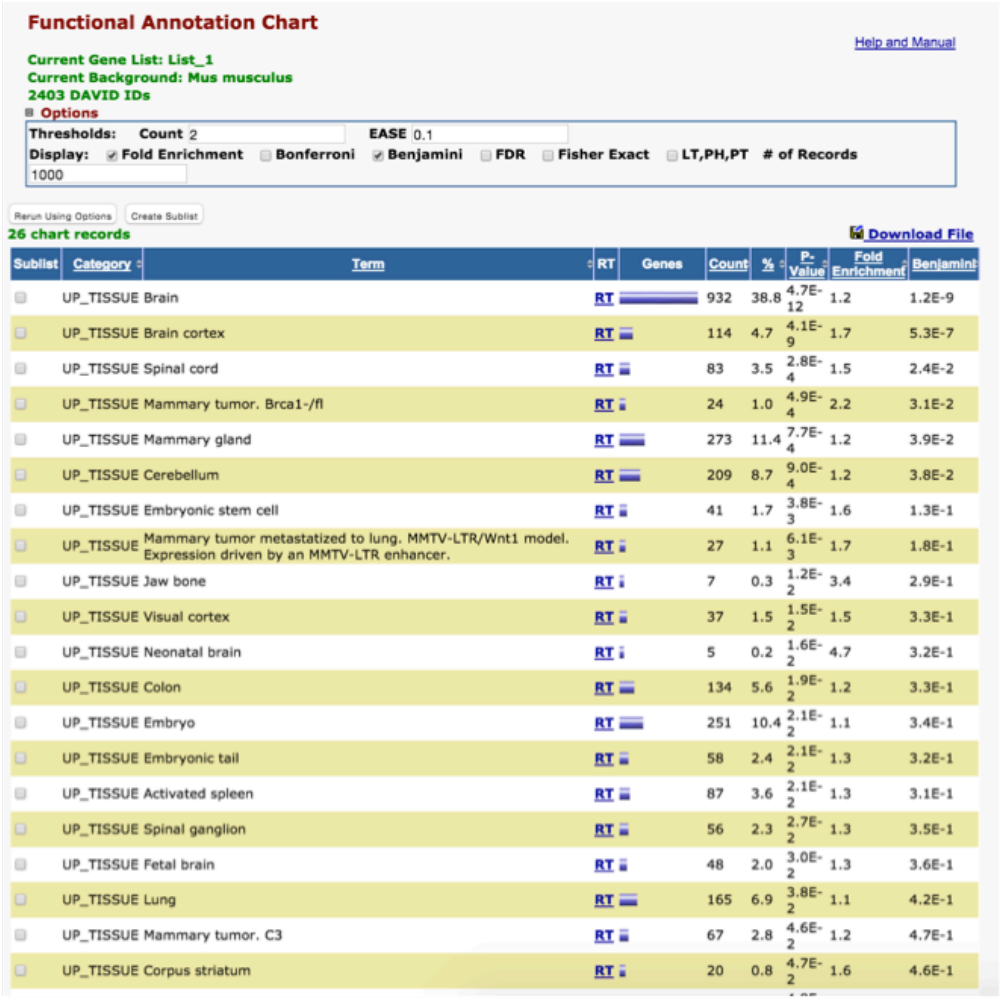
Supplementary Table 4: Top 10 most statistically significant hypo-DMRs identified between mNSCs and mESCs. Associated genes refer to the nearest gene falling within 100kb upstream and 50kb downstream of the DMR. DMRs falling within repeat regions, with a CpG count of zero, were excluded from the list. Gene annotations are taken from the NCBI Gene database and the MGI database.

Appendix 2: Supplementary Figures



Supplementary Figure 1: A) Primer pair designed around the Grm4-associated hyper-hDMR (chr17:27,422,334-27,422,558). B) Primer pair designed around the Jarid2-associated hyper-DMR (chr13:44,705,585-44,705,763).

A



Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Mus musculus
2403 DAVID IDs

Options

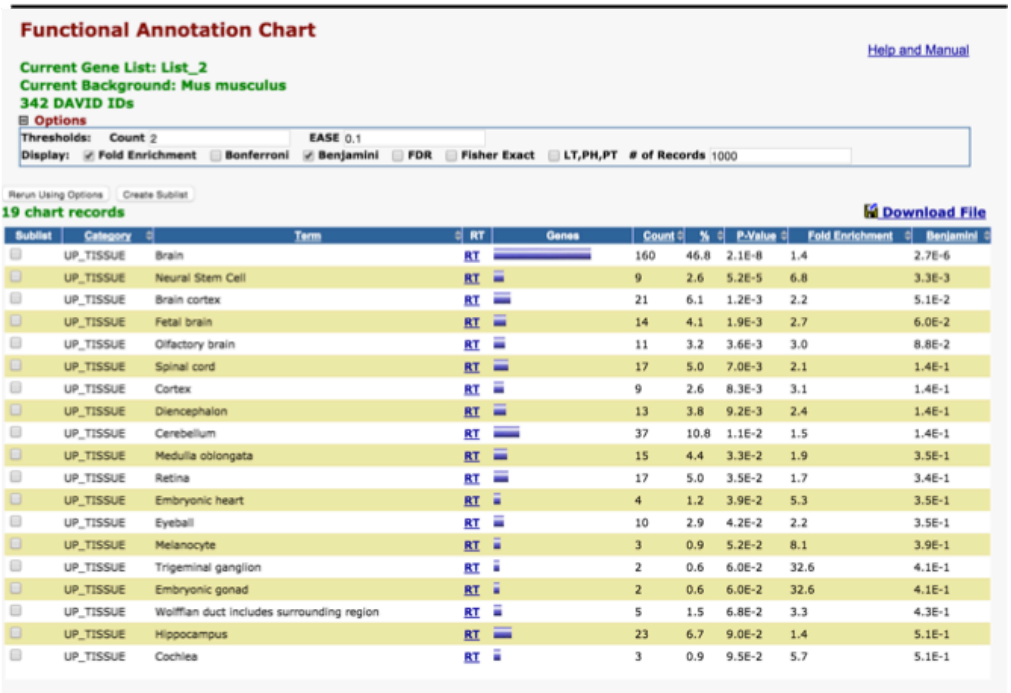
Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment Bonferroni Benjamini FDR Fisher Exact LT,PH,PT # of Records 1000

Rerun Using Options | Create Sublist

26 chart records — Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	UP_TISSUE	Brain	RT		932	38.8	4.7E-12	1.2	1.2E-9
	UP_TISSUE	Brain cortex	RT		114	4.7	4.1E-9	1.7	5.3E-7
	UP_TISSUE	Spinal cord	RT		83	3.5	2.8E-4	1.5	2.4E-2
	UP_TISSUE	Mammary tumor. Brca1-/fl	RT		24	1.0	4.9E-4	2.2	3.1E-2
	UP_TISSUE	Mammary gland	RT		273	11.4	7.7E-4	1.2	3.9E-2
	UP_TISSUE	Cerebellum	RT		209	8.7	9.0E-4	1.2	3.8E-2
	UP_TISSUE	Embryonic stem cell	RT		41	1.7	3.8E-3	1.6	1.3E-1
	UP_TISSUE	Mammary tumor metastatized to lung. MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer.	RT		27	1.1	6.1E-3	1.7	1.8E-1
	UP_TISSUE	Jaw bone	RT		7	0.3	1.2E-2	3.4	2.9E-1
	UP_TISSUE	Visual cortex	RT		37	1.5	1.5E-2	1.5	3.3E-1
	UP_TISSUE	Neonatal brain	RT		5	0.2	1.6E-2	4.7	3.2E-1
	UP_TISSUE	Colon	RT		134	5.6	1.9E-2	1.2	3.3E-1
	UP_TISSUE	Embryo	RT		251	10.4	2.1E-2	1.1	3.4E-1
	UP_TISSUE	Embryonic tail	RT		58	2.4	2.1E-2	1.3	3.2E-1
	UP_TISSUE	Activated spleen	RT		87	3.6	2.1E-2	1.3	3.1E-1
	UP_TISSUE	Spinal ganglion	RT		56	2.3	2.7E-2	1.3	3.5E-1
	UP_TISSUE	Fetal brain	RT		48	2.0	3.0E-2	1.3	3.6E-1
	UP_TISSUE	Lung	RT		165	6.9	3.8E-2	1.1	4.2E-1
	UP_TISSUE	Mammary tumor. C3	RT		67	2.8	4.6E-2	1.2	4.7E-1
	UP_TISSUE	Corpus striatum	RT		20	0.8	4.7E-2	1.6	4.6E-1

B



Functional Annotation Chart

Help and Manual

Current Gene List: List_2
Current Background: Mus musculus
342 DAVID IDs

Options

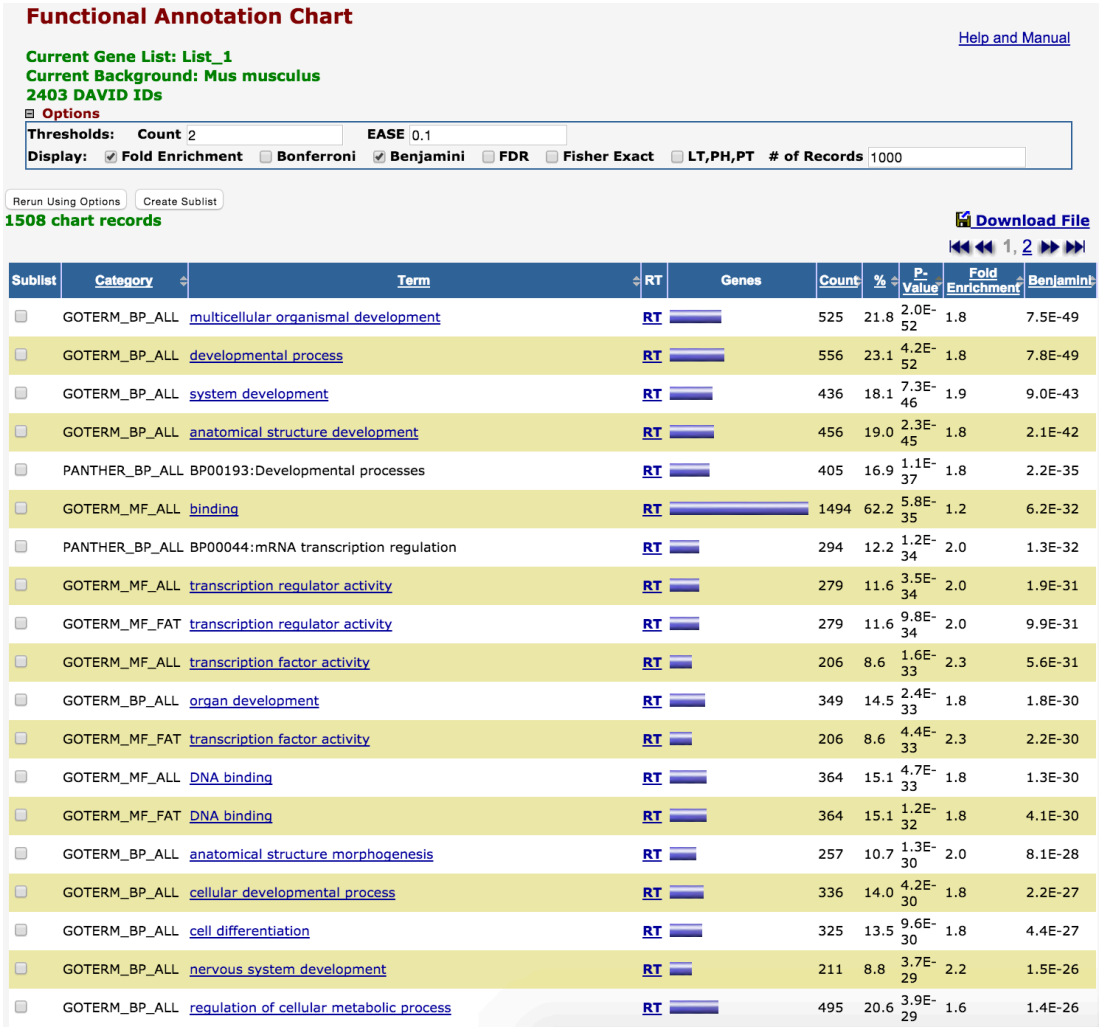
Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment Bonferroni Benjamini FDR Fisher Exact LT,PH,PT # of Records 1000

Rerun Using Options | Create Sublist

19 chart records — Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	UP_TISSUE	Brain	RT		160	46.8	2.1E-8	1.4	2.7E-6
	UP_TISSUE	Neural Stem Cell	RT		9	2.6	5.2E-5	6.8	3.3E-3
	UP_TISSUE	Brain cortex	RT		21	6.1	1.2E-3	2.2	5.1E-2
	UP_TISSUE	Fetal brain	RT		14	4.1	1.9E-3	2.7	6.0E-2
	UP_TISSUE	Olfactory brain	RT		11	3.2	3.6E-3	3.0	8.8E-2
	UP_TISSUE	Spinal cord	RT		17	5.0	7.0E-3	2.1	1.4E-1
	UP_TISSUE	Cortex	RT		9	2.6	8.3E-3	3.1	1.4E-1
	UP_TISSUE	Diencephalon	RT		13	3.8	9.2E-3	2.4	1.4E-1
	UP_TISSUE	Cerebellum	RT		37	10.8	1.1E-2	1.5	1.4E-1
	UP_TISSUE	Medulla oblongata	RT		15	4.4	3.3E-2	1.9	3.5E-1
	UP_TISSUE	Retina	RT		17	5.0	3.5E-2	1.7	3.4E-1
	UP_TISSUE	Embryonic heart	RT		4	1.2	3.9E-2	5.3	3.5E-1
	UP_TISSUE	Eyeball	RT		10	2.9	4.2E-2	2.2	3.5E-1
	UP_TISSUE	Melanocyte	RT		3	0.9	5.2E-2	8.1	3.9E-1
	UP_TISSUE	Trigeminal ganglion	RT		2	0.6	6.0E-2	32.6	4.1E-1
	UP_TISSUE	Embryonic gonad	RT		2	0.6	6.0E-2	32.6	4.1E-1
	UP_TISSUE	Wolffian duct includes surrounding region	RT		5	1.5	6.8E-2	3.3	4.3E-1
	UP_TISSUE	Hippocampus	RT		23	6.7	9.0E-2	1.4	5.1E-1
	UP_TISSUE	Cochlea	RT		3	0.9	9.5E-2	5.7	5.1E-1

Supplementary Figure 2: DAVID functional annotation of both hyper-DMR (A) and hypo-DMR (B) gene-sets identified during the mESC to mNSC transition. Gene enrichment scores for tissue expression terms are above. Both gene-sets show the most significant enrichment for genes upregulated in brain tissue.



Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Mus musculus
2403 DAVID IDs

Options

Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment Bonferroni Benjamini FDR Fisher Exact LT,PH,PT # of Records 1000

Rerun Using Options | Create Sublist

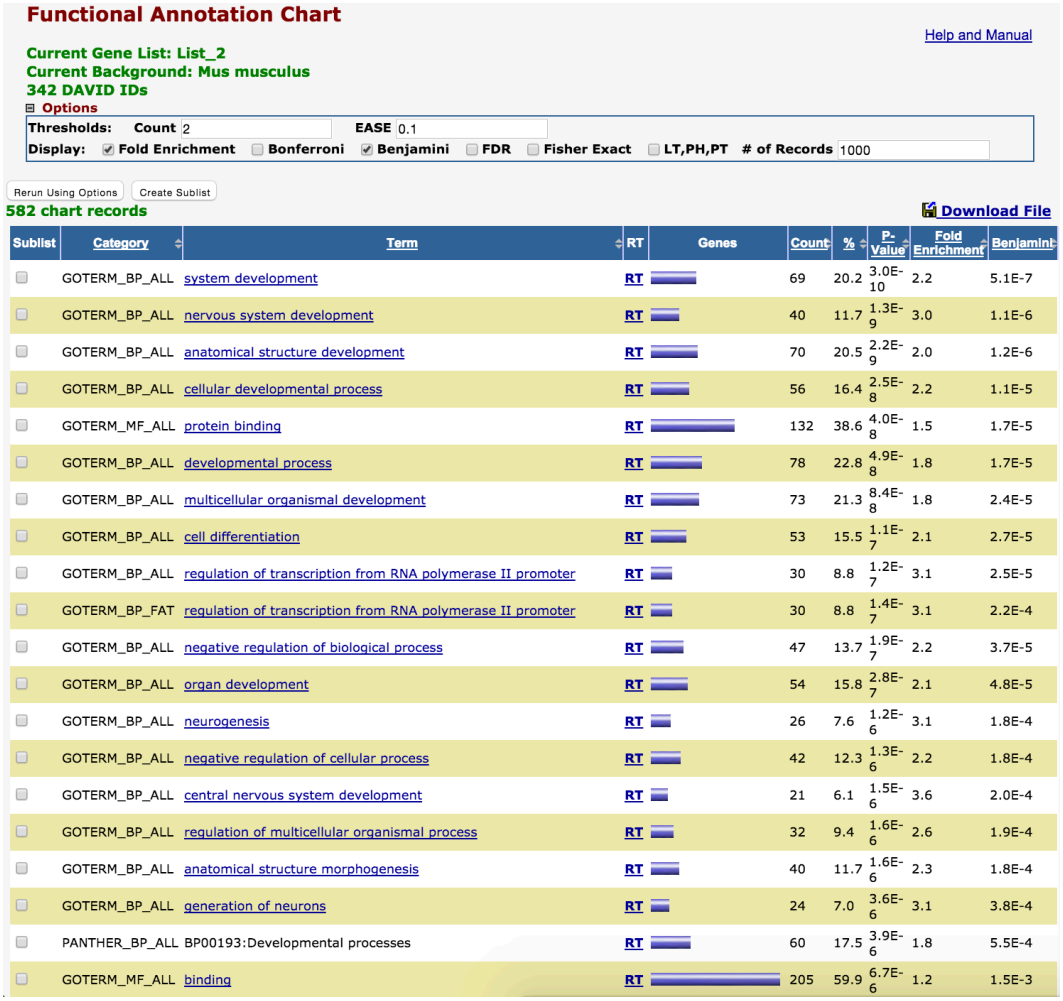
1508 chart records

Download File

1, 2

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	GOTERM_BP_ALL	multicellular organismal development	RT		525	21.8	2.0E-52	1.8	7.5E-49
	GOTERM_BP_ALL	developmental process	RT		556	23.1	4.2E-52	1.8	7.8E-49
	GOTERM_BP_ALL	system development	RT		436	18.1	7.3E-46	1.9	9.0E-43
	GOTERM_BP_ALL	anatomical structure development	RT		456	19.0	2.3E-45	1.8	2.1E-42
	PANTHER_BP_ALL	BP00193:Developmental processes	RT		405	16.9	1.1E-37	1.8	2.2E-35
	GOTERM_MF_ALL	binding	RT		1494	62.2	5.8E-35	1.2	6.2E-32
	PANTHER_BP_ALL	BP00044:mRNA transcription regulation	RT		294	12.2	1.2E-34	2.0	1.3E-32
	GOTERM_MF_ALL	transcription regulator activity	RT		279	11.6	3.5E-34	2.0	1.9E-31
	GOTERM_MF_FAT	transcription regulator activity	RT		279	11.6	9.8E-34	2.0	9.9E-31
	GOTERM_MF_ALL	transcription factor activity	RT		206	8.6	1.6E-33	2.3	5.6E-31
	GOTERM_BP_ALL	organ development	RT		349	14.5	2.4E-33	1.8	1.8E-30
	GOTERM_MF_FAT	transcription factor activity	RT		206	8.6	4.4E-33	2.3	2.2E-30
	GOTERM_MF_ALL	DNA binding	RT		364	15.1	4.7E-33	1.8	1.3E-30
	GOTERM_MF_FAT	DNA binding	RT		364	15.1	1.2E-32	1.8	4.1E-30
	GOTERM_BP_ALL	anatomical structure morphogenesis	RT		257	10.7	1.3E-30	2.0	8.1E-28
	GOTERM_BP_ALL	cellular developmental process	RT		336	14.0	4.2E-30	1.8	2.2E-27
	GOTERM_BP_ALL	cell differentiation	RT		325	13.5	9.6E-30	1.8	4.4E-27
	GOTERM_BP_ALL	nervous system development	RT		211	8.8	3.7E-29	2.2	1.5E-26
	GOTERM_BP_ALL	regulation of cellular metabolic process	RT		495	20.6	3.9E-29	1.6	1.4E-26

Supplementary Figure 3: DAVID gene ontology (GO) of the hyper-DMR gene-set corresponding to the mESC to mNSC transition. Gene enrichment scores for GO terms are shown above. The most significantly enriched term was multicellular organismal development, although other related terms such as organ development, transcriptional regulation, cell differentiation and nervous system development were highly enriched for genes associated with hyper-DMRs and occurred within the top 20 most significant GO terms.



Functional Annotation Chart

Help and Manual

Current Gene List: List_2
Current Background: Mus musculus
342 DAVID IDs

Options

Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment | Bonferroni | Benjamini | FDR | Fisher Exact | LT,PH,PT | # of Records 1000

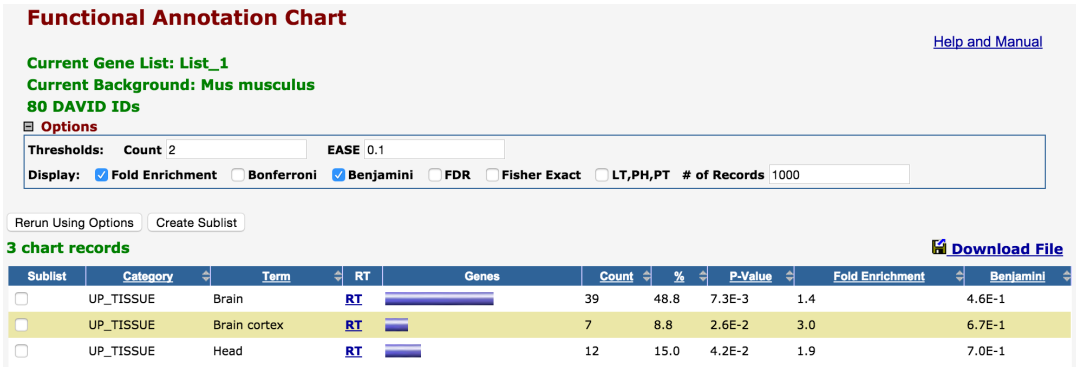
Rerun Using Options | Create Sublist

582 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	GOTERM_BP_ALL	system development	RT		69	20.2	3.0E-10	2.2	5.1E-7
	GOTERM_BP_ALL	nervous system development	RT		40	11.7	1.3E-9	3.0	1.1E-6
	GOTERM_BP_ALL	anatomical structure development	RT		70	20.5	2.2E-9	2.0	1.2E-6
	GOTERM_BP_ALL	cellular developmental process	RT		56	16.4	2.5E-8	2.2	1.1E-5
	GOTERM_MF_ALL	protein binding	RT		132	38.6	4.0E-8	1.5	1.7E-5
	GOTERM_BP_ALL	developmental process	RT		78	22.8	4.9E-8	1.8	1.7E-5
	GOTERM_BP_ALL	multicellular organismal development	RT		73	21.3	8.4E-8	1.8	2.4E-5
	GOTERM_BP_ALL	cell differentiation	RT		53	15.5	1.1E-7	2.1	2.7E-5
	GOTERM_BP_ALL	regulation of transcription from RNA polymerase II promoter	RT		30	8.8	1.2E-7	3.1	2.5E-5
	GOTERM_BP_FAT	regulation of transcription from RNA polymerase II promoter	RT		30	8.8	1.4E-7	3.1	2.2E-4
	GOTERM_BP_ALL	negative regulation of biological process	RT		47	13.7	1.9E-7	2.2	3.7E-5
	GOTERM_BP_ALL	organ development	RT		54	15.8	2.8E-7	2.1	4.8E-5
	GOTERM_BP_ALL	neurogenesis	RT		26	7.6	1.2E-6	3.1	1.8E-4
	GOTERM_BP_ALL	negative regulation of cellular process	RT		42	12.3	1.3E-6	2.2	1.8E-4
	GOTERM_BP_ALL	central nervous system development	RT		21	6.1	1.5E-6	3.6	2.0E-4
	GOTERM_BP_ALL	regulation of multicellular organismal process	RT		32	9.4	1.6E-6	2.6	1.9E-4
	GOTERM_BP_ALL	anatomical structure morphogenesis	RT		40	11.7	1.6E-6	2.3	1.8E-4
	GOTERM_BP_ALL	generation of neurons	RT		24	7.0	3.6E-6	3.1	3.8E-4
	PANTHER_BP_ALL	BP00193:Developmental processes	RT		60	17.5	3.9E-6	1.8	5.5E-4
	GOTERM_MF_ALL	binding	RT		205	59.9	6.7E-6	1.2	1.5E-3

Supplementary Figure 4: DAVID gene ontology (GO) of the hypo-DMR gene-set corresponding to the mESC to mNSC transition. Gene enrichment scores for GO terms are shown above. The most significantly enriched term was system development, closely followed by nervous system development and anatomical structure development. Other highly enriched GO terms occurring within the top 20 include cell differentiation, neurogenesis, central nervous system development and generation of neurons.



Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Mus musculus
80 DAVID IDs

Options

Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment | Bonferroni | Benjamini | FDR | Fisher Exact | LT,PH,PT | # of Records 1000

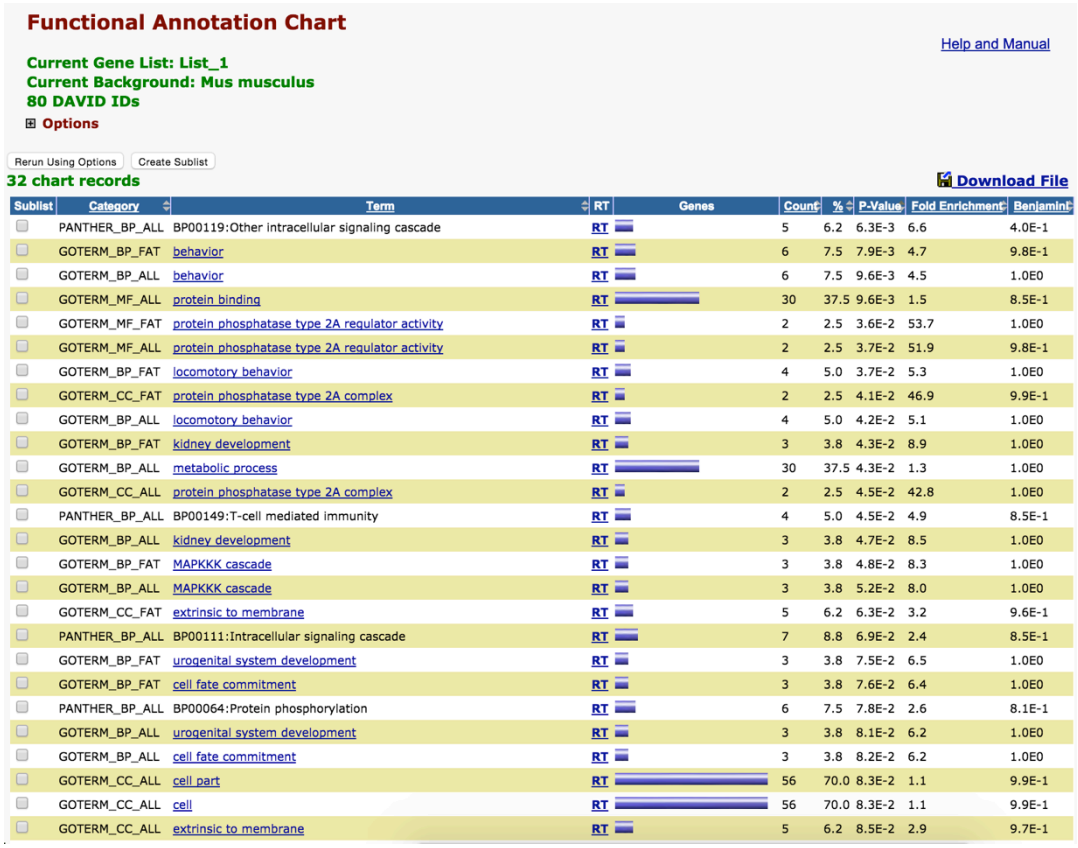
Rerun Using Options | Create Sublist

3 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	UP_TISSUE	Brain	RT		39	48.8	7.3E-3	1.4	4.6E-1
	UP_TISSUE	Brain cortex	RT		7	8.8	2.6E-2	3.0	6.7E-1
	UP_TISSUE	Head	RT		12	15.0	4.2E-2	1.9	7.0E-1

Supplementary Figure 5: DAVID functional annotation of the hyper-DMR gene-set identified during the mNSC to astrocyte transition. Gene enrichment scores for the annotation category "tissue expression" are shown. Only 3 terms (brain, brain cortex and head) showed significant enrichment.



Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Mus musculus
80 DAVID IDs

⊞ Options

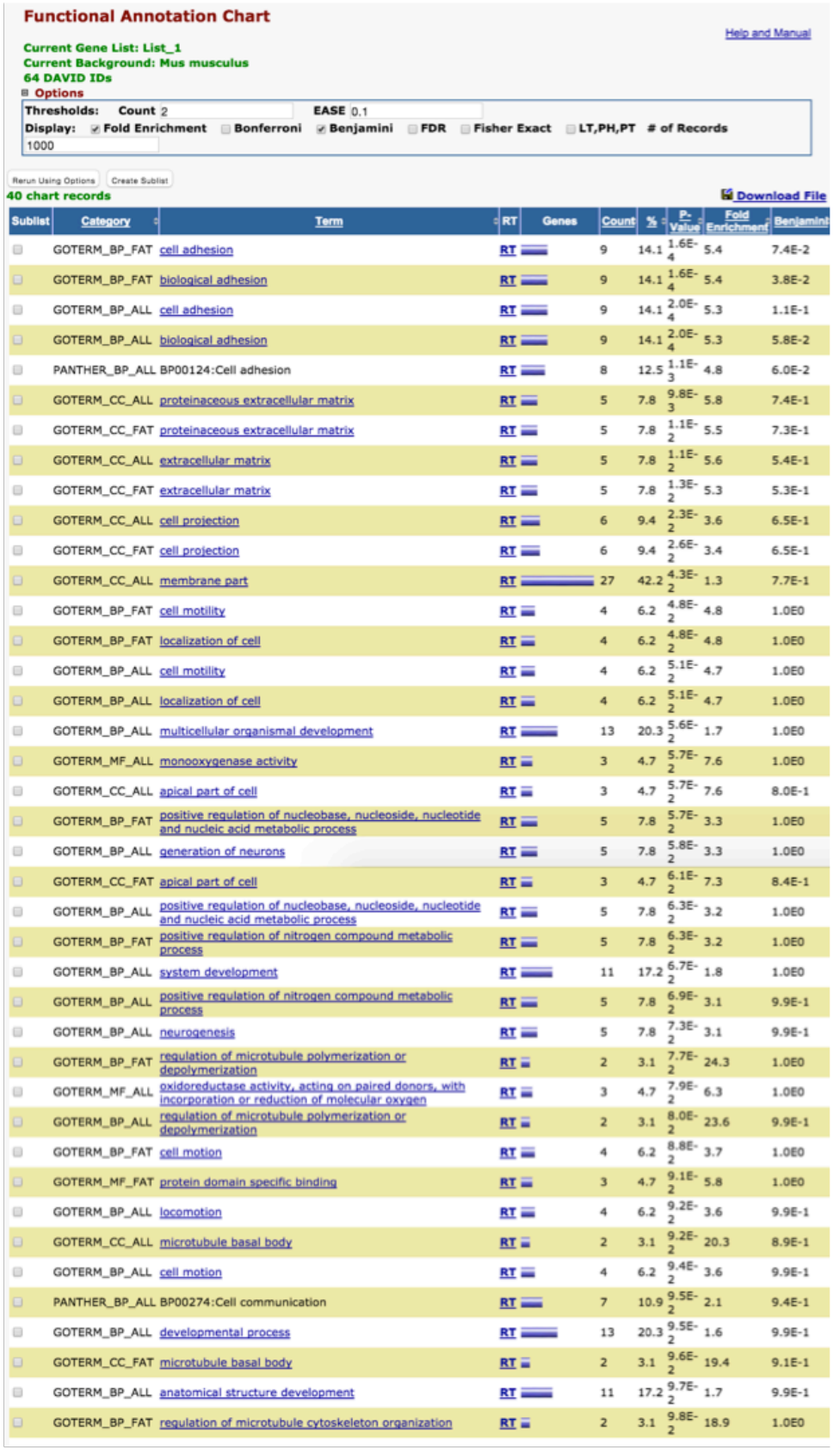
Rerun Using Options | Create Sublist

32 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	PANTHER_BP_ALL	BP00119:Other intracellular signaling cascade	RT		5	6.2	6.3E-3	6.6	4.0E-1
	GOTERM_BP_FAT	behavior	RT		6	7.5	7.9E-3	4.7	9.8E-1
	GOTERM_BP_ALL	behavior	RT		6	7.5	9.6E-3	4.5	1.0E0
	GOTERM_MF_ALL	protein binding	RT		30	37.5	9.6E-3	1.5	8.5E-1
	GOTERM_MF_FAT	protein phosphatase type 2A regulator activity	RT		2	2.5	3.6E-2	53.7	1.0E0
	GOTERM_MF_ALL	protein phosphatase type 2A regulator activity	RT		2	2.5	3.7E-2	51.9	9.8E-1
	GOTERM_BP_FAT	locomotory behavior	RT		4	5.0	3.7E-2	5.3	1.0E0
	GOTERM_CC_FAT	protein phosphatase type 2A complex	RT		2	2.5	4.1E-2	46.9	9.9E-1
	GOTERM_BP_ALL	locomotory behavior	RT		4	5.0	4.2E-2	5.1	1.0E0
	GOTERM_BP_FAT	kidney development	RT		3	3.8	4.3E-2	8.9	1.0E0
	GOTERM_BP_ALL	metabolic process	RT		30	37.5	4.3E-2	1.3	1.0E0
	GOTERM_CC_ALL	protein phosphatase type 2A complex	RT		2	2.5	4.5E-2	42.8	1.0E0
	PANTHER_BP_ALL	BP00149:T-cell mediated immunity	RT		4	5.0	4.5E-2	4.9	8.5E-1
	GOTERM_BP_ALL	kidney development	RT		3	3.8	4.7E-2	8.5	1.0E0
	GOTERM_BP_FAT	MAPKKK cascade	RT		3	3.8	4.8E-2	8.3	1.0E0
	GOTERM_BP_ALL	MAPKKK cascade	RT		3	3.8	5.2E-2	8.0	1.0E0
	GOTERM_CC_FAT	extrinsic to membrane	RT		5	6.2	6.3E-2	3.2	9.6E-1
	PANTHER_BP_ALL	BP00111:Intracellular signaling cascade	RT		7	8.8	6.9E-2	2.4	8.5E-1
	GOTERM_BP_FAT	urogenital system development	RT		3	3.8	7.5E-2	6.5	1.0E0
	GOTERM_BP_FAT	cell fate commitment	RT		3	3.8	7.6E-2	6.4	1.0E0
	PANTHER_BP_ALL	BP00064:Protein phosphorylation	RT		6	7.5	7.8E-2	2.6	8.1E-1
	GOTERM_BP_ALL	urogenital system development	RT		3	3.8	8.1E-2	6.2	1.0E0
	GOTERM_BP_ALL	cell fate commitment	RT		3	3.8	8.2E-2	6.2	1.0E0
	GOTERM_CC_ALL	cell part	RT		56	70.0	8.3E-2	1.1	9.9E-1
	GOTERM_CC_ALL	cell	RT		56	70.0	8.3E-2	1.1	9.9E-1
	GOTERM_CC_ALL	extrinsic to membrane	RT		5	6.2	8.5E-2	2.9	9.7E-1

Supplementary Figure 6: DAVID GO analysis of the hyper-DMR gene-set identified during the mNSC to astrocyte transition. Gene enrichment scores for GO terms are shown above.



Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Mus musculus
64 DAVID IDs

Options

Thresholds: Count 2 EASE 0.1
Display: Fold Enrichment Bonferroni Benjamini FDR Fisher Exact LT,PH,PT # of Records
1000

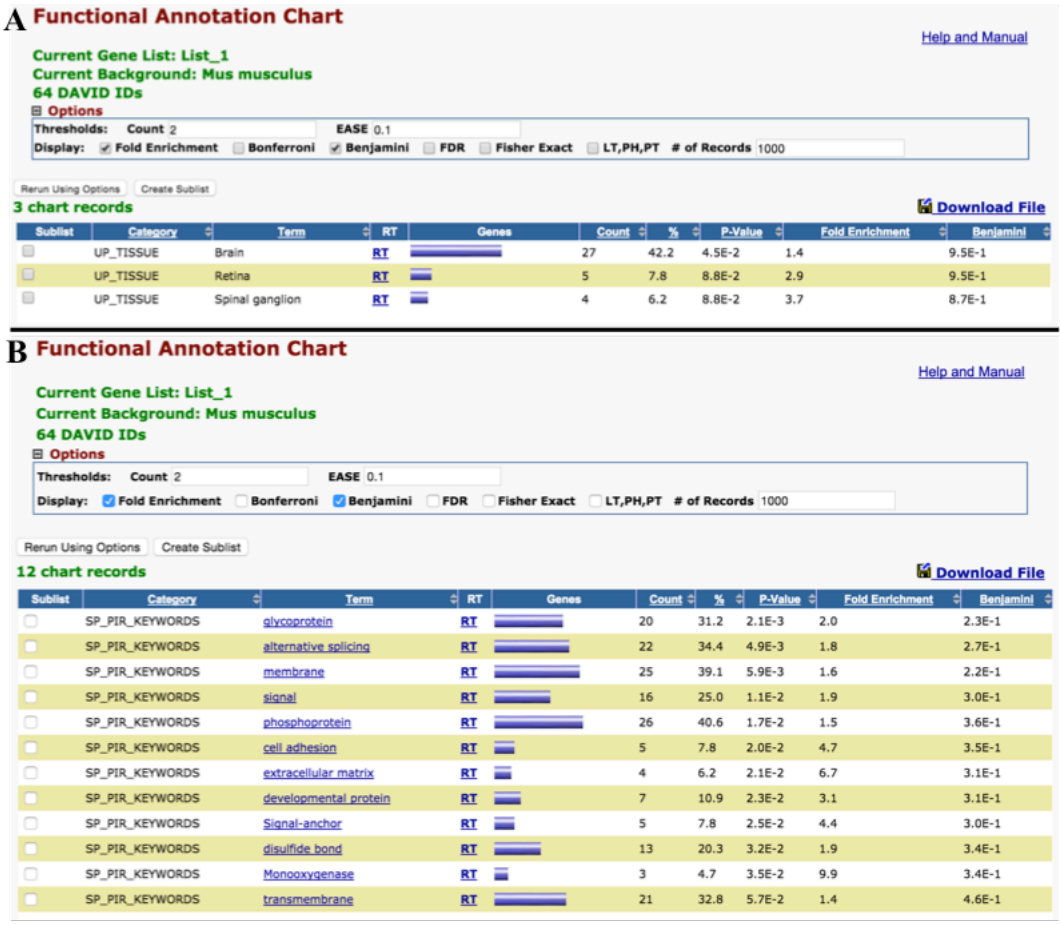
Rerun Using Options | Create Sublist

40 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Fold Enrichment	Benjamini
	GOTERM_BP_FAT	cell adhesion	RT		9	14.1	1.6E-4	5.4	7.4E-2
	GOTERM_BP_FAT	biological adhesion	RT		9	14.1	1.6E-4	5.4	3.8E-2
	GOTERM_BP_ALL	cell adhesion	RT		9	14.1	2.0E-4	5.3	1.1E-1
	GOTERM_BP_ALL	biological adhesion	RT		9	14.1	2.0E-4	5.3	5.8E-2
	PANTHER_BP_ALL	BP00124:Cell adhesion	RT		8	12.5	1.1E-3	4.8	6.0E-2
	GOTERM_CC_ALL	proteinaceous extracellular matrix	RT		5	7.8	9.8E-3	5.8	7.4E-1
	GOTERM_CC_FAT	proteinaceous extracellular matrix	RT		5	7.8	1.1E-2	5.5	7.3E-1
	GOTERM_CC_ALL	extracellular matrix	RT		5	7.8	1.1E-2	5.6	5.4E-1
	GOTERM_CC_FAT	extracellular matrix	RT		5	7.8	1.3E-2	5.3	5.3E-1
	GOTERM_CC_ALL	cell projection	RT		6	9.4	2.3E-2	3.6	6.5E-1
	GOTERM_CC_FAT	cell projection	RT		6	9.4	2.6E-2	3.4	6.5E-1
	GOTERM_CC_ALL	membrane part	RT		27	42.2	4.3E-2	1.3	7.7E-1
	GOTERM_BP_FAT	cell motility	RT		4	6.2	4.8E-2	4.8	1.0E0
	GOTERM_BP_FAT	localization of cell	RT		4	6.2	4.8E-2	4.8	1.0E0
	GOTERM_BP_ALL	cell motility	RT		4	6.2	5.1E-2	4.7	1.0E0
	GOTERM_BP_ALL	localization of cell	RT		4	6.2	5.1E-2	4.7	1.0E0
	GOTERM_BP_ALL	multicellular organismal development	RT		13	20.3	5.6E-2	1.7	1.0E0
	GOTERM_MF_ALL	monooxygenase activity	RT		3	4.7	5.7E-2	7.6	1.0E0
	GOTERM_CC_ALL	apical part of cell	RT		3	4.7	5.7E-2	7.6	8.0E-1
	GOTERM_BP_FAT	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		5	7.8	5.7E-2	3.3	1.0E0
	GOTERM_BP_ALL	generation of neurons	RT		5	7.8	5.8E-2	3.3	1.0E0
	GOTERM_CC_FAT	apical part of cell	RT		3	4.7	6.1E-2	7.3	8.4E-1
	GOTERM_BP_ALL	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		5	7.8	6.3E-2	3.2	1.0E0
	GOTERM_BP_FAT	positive regulation of nitrogen compound metabolic process	RT		5	7.8	6.3E-2	3.2	1.0E0
	GOTERM_BP_ALL	system development	RT		11	17.2	6.7E-2	1.8	1.0E0
	GOTERM_BP_ALL	positive regulation of nitrogen compound metabolic process	RT		5	7.8	6.9E-2	3.1	9.9E-1
	GOTERM_BP_ALL	neurogenesis	RT		5	7.8	7.3E-2	3.1	9.9E-1
	GOTERM_BP_FAT	regulation of microtubule polymerization or depolymerization	RT		2	3.1	7.7E-2	24.3	1.0E0
	GOTERM_MF_ALL	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	RT		3	4.7	7.9E-2	6.3	1.0E0
	GOTERM_BP_ALL	regulation of microtubule polymerization or depolymerization	RT		2	3.1	8.0E-2	23.6	9.9E-1
	GOTERM_BP_FAT	cell motion	RT		4	6.2	8.8E-2	3.7	1.0E0
	GOTERM_MF_FAT	protein domain specific binding	RT		3	4.7	9.1E-2	5.8	1.0E0
	GOTERM_BP_ALL	locomotion	RT		4	6.2	9.2E-2	3.6	9.9E-1
	GOTERM_CC_ALL	microtubule basal body	RT		2	3.1	9.2E-2	20.3	8.9E-1
	GOTERM_BP_ALL	cell motion	RT		4	6.2	9.4E-2	3.6	9.9E-1
	PANTHER_BP_ALL	BP00274:Cell communication	RT		7	10.9	9.5E-2	2.1	9.4E-1
	GOTERM_BP_ALL	developmental process	RT		13	20.3	9.5E-2	1.6	9.9E-1
	GOTERM_CC_FAT	microtubule basal body	RT		2	3.1	9.6E-2	19.4	9.1E-1
	GOTERM_BP_ALL	anatomical structure development	RT		11	17.2	9.7E-2	1.7	9.9E-1
	GOTERM_BP_FAT	regulation of microtubule cytoskeleton organization	RT		2	3.1	9.8E-2	18.9	1.0E0

Supplementary Figure 7: DAVID analysis of the hyper-hDMR gene set identified between mESCs and mNSCs. All 40 gene ontology terms enriched for hDMR-associated genes are detailed above.



Supplementary Figure 8: DAVID analysis of the hyper-hDMR associated gene set identified between mESCs and mNSCs. All terms enriched for hDMR-associated genes falling within either (A) tissue expression or (B) Swiss-Prot Protein Information Resource (SP-PIR) keyword categories are detailed above.

MGI Expression: Detected (7 terms)

Table controls: Export | Shown top rows in this table: 20 Set | Term annotation count: Min: 1 Max: Inf Set | Visualize this table: [select one]

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
TS23_palatal shelf	20	8.6021e-7	4.0353e-4	3.6547	20	8.55%	6	4.7520e-2	6.6099	8	141	4.26%
TS23_upper arm epidermis	23	2.1302e-6	8.6894e-4	25.0974	5	2.14%	1	5.7663e-2	69.9000	3	5	1.60%
TS23_upper arm skin	25	3.0659e-6	1.1506e-3	23.2757	5	2.14%	4	2.8649e-2	58.2500	3	6	1.60%
TS23_forearm skin	26	3.1056e-6	1.1207e-3	23.2137	5	2.14%	1	5.7663e-2	69.9000	3	5	1.60%
TS17_1st branchial arch maxillary component endoderm	28	1.0894e-5	3.6503e-3	17.8731	5	2.14%	1	5.7663e-2	69.9000	3	5	1.60%
TS19_myelencephalon ventricular layer	31	2.8285e-5	8.5604e-3	14.6199	5	2.14%	4	2.8649e-2	58.2500	3	6	1.60%
TS17_1st branchial arch mandibular component endoderm	34	3.2022e-5	8.8363e-3	14.2404	5	2.14%	7	4.5260e-2	43.6875	3	8	1.60%

GO Biological Process (4 terms)

Table controls: Export | Shown top rows in this table: 20 Set | Term annotation count: Min: 1 Max: Inf Set | Visualize this table: [select one]

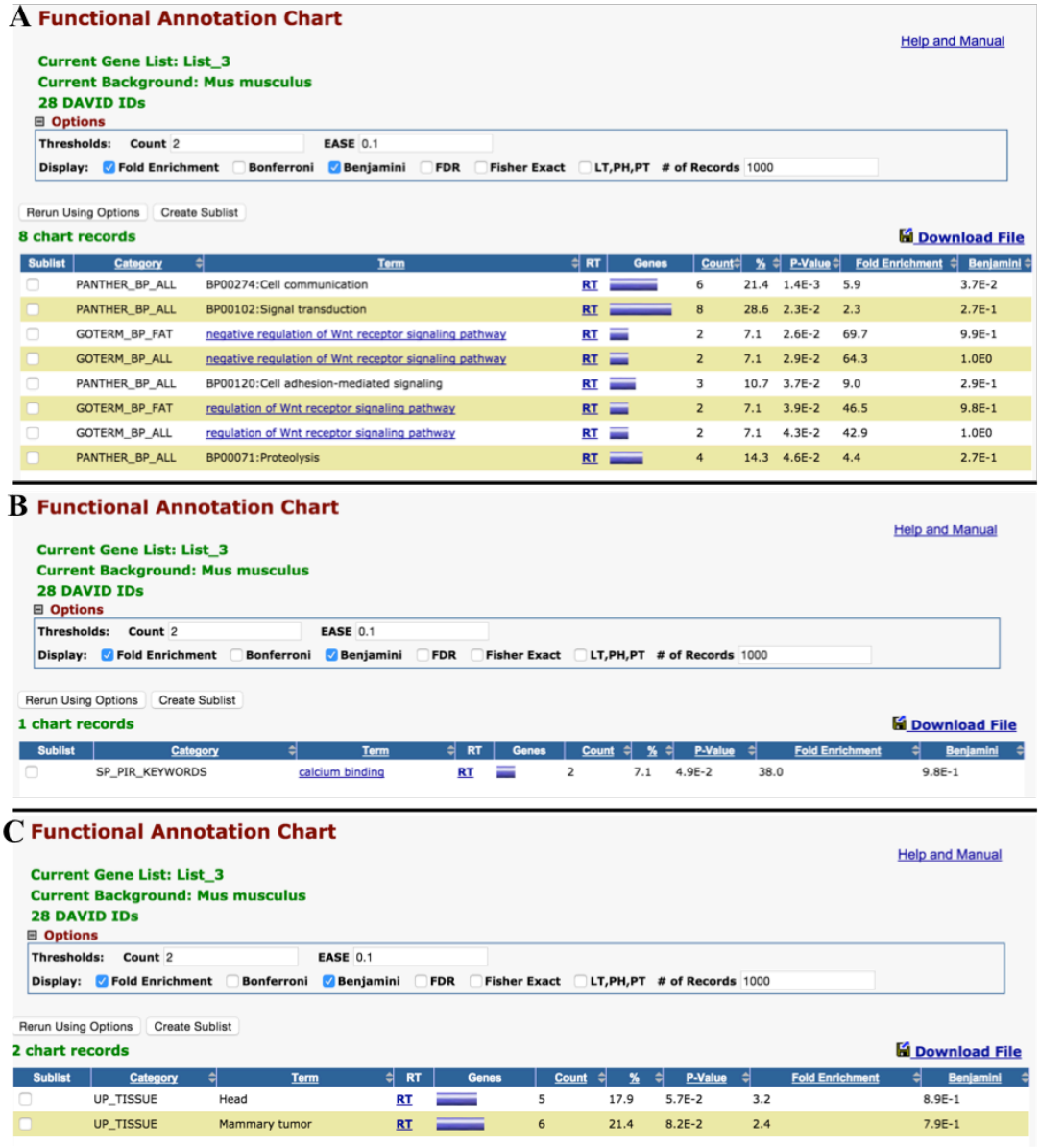
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
respiratory electron transport chain	3	1.5098e-17	5.0670e-14	34.9105	14	5.98%	2	2.9228e-3	19.4167	6	36	3.19%
oxidative phosphorylation	4	5.5229e-17	1.3901e-13	65.0293	11	4.70%	3	1.2413e-2	20.8036	5	28	2.66%
electron transport chain	5	1.4448e-16	2.9093e-13	29.5749	14	5.98%	1	5.4296e-4	19.8902	7	41	3.72%
cellular respiration	10	2.9924e-10	3.0127e-7	9.8632	14	5.98%	4	4.2142e-2	8.6755	7	94	3.72%

GO Molecular Function (5 terms)

Table controls: Export | Shown top rows in this table: 20 Set | Term annotation count: Min: 1 Max: Inf Set | Visualize this table: [select one]

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
NADH dehydrogenase (ubiquinone) activity	2	1.1136e-20	1.9427e-17	59.0401	14	5.98%	1	4.5177e-6	32.6200	7	25	3.72%
oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	3	9.1776e-19	1.0674e-15	42.8399	14	5.98%	2	9.2258e-6	27.1833	7	30	3.72%
oxidoreductase activity, acting on NAD(P)H	5	4.9929e-13	3.4840e-10	16.1220	14	5.98%	3	1.6327e-3	12.5462	7	65	3.72%
N,N-dimethylaniline monooxygenase activity	23	9.4027e-5	1.4263e-2	35.4046	3	1.28%	5	3.5122e-2	38.8333	3	9	1.60%
oxidoreductase activity	26	2.1917e-4	2.9410e-2	2.1941	25	10.68%	4	8.5005e-3	3.0537	20	763	10.64%

Supplementary Figure 9: GREAT functional annotation of hyper-hDMR regions identified between mESCs and mNSCs. hDMR-associated genes were enriched for GO terms classified under mouse tissue expression, biological processes and molecular function categories, as detailed above.



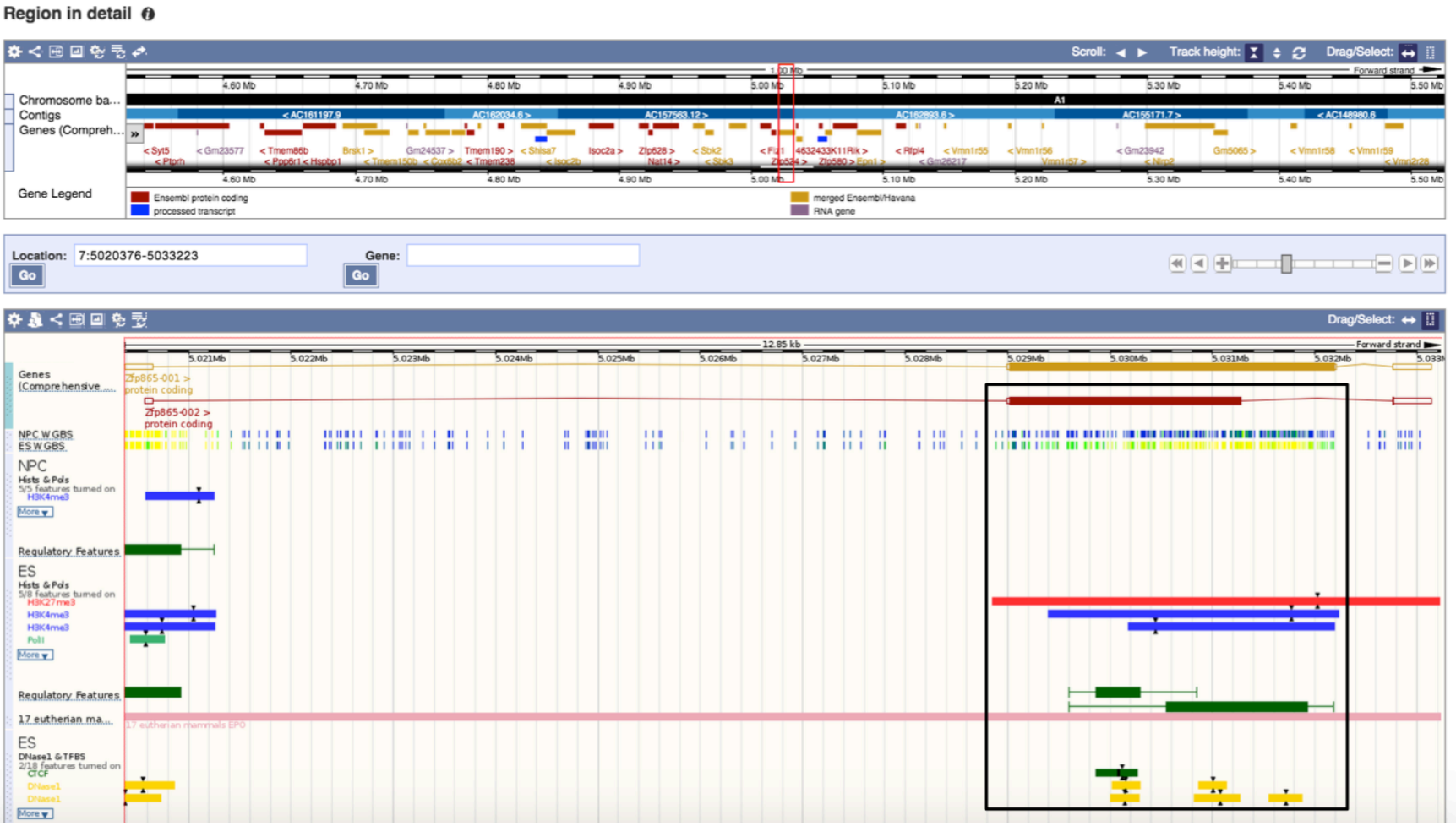
Supplementary Figure 10: DAVID analysis of the hypo-hDMR gene set identified between mESCs and mNSCs. Gene enrichments for (A) gene ontology, (B) PIR keywords or (C) tissue expression terms are detailed above.

MGI Expression: Detected (9 terms) — Global controls

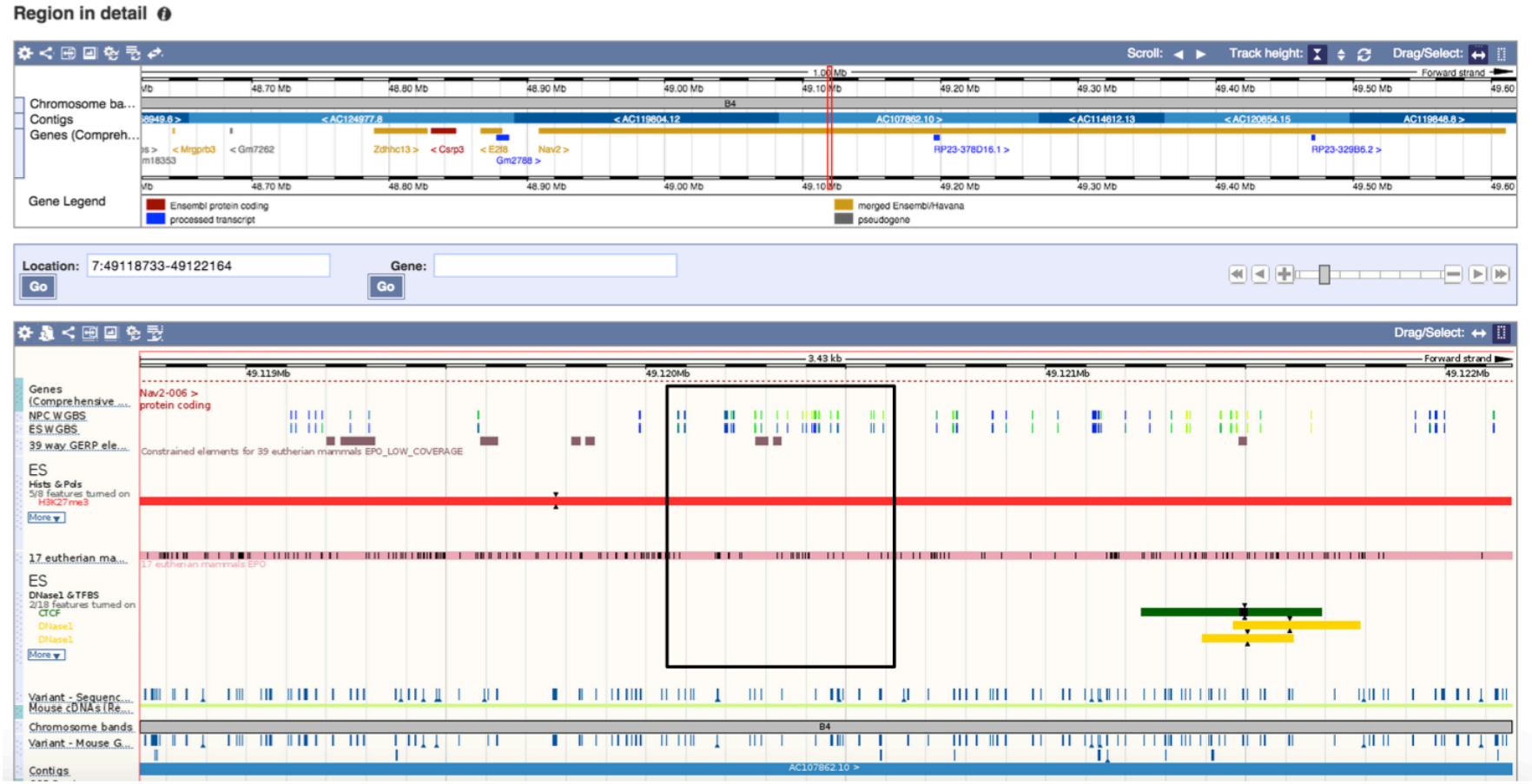
Table controls: Export | Shown top rows in this table: 20 Set | Term annotation count: Min: 1 Max: Inf Set | Visualize this table: NEW [select one]

Term Name	Binom Rank	Binom Raw P-Value	*Binom FDR Q-Val*	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	*Hyper FDR Q-Val*	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
TS17_branchial arch	1	2.8885e-6	2.7100e-2	2.3183	34	21.66%	4	8.9258e-5	3.4643	26	761	12.04%
TS19_1st branchial arch	2	4.0533e-6	1.9014e-2	3.7111	17	10.83%	7	1.2853e-4	6.6240	13	199	6.02%
TS19_oral region	3	4.2489e-6	1.3288e-2	3.0117	22	14.01%	1	4.5390e-6	6.0205	19	320	8.80%
TS19_branchial arch	4	6.3236e-6	1.4832e-2	3.0370	21	13.38%	6	1.0990e-4	5.3019	16	306	7.41%
TS19_2nd branchial arch	5	1.0650e-5	1.9984e-2	4.0793	14	8.92%	12	1.1087e-3	7.1914	10	141	4.63%
TS28_hippocampus granule cell layer	6	2.0637e-5	3.2269e-2	15.5659	5	3.18%	142	2.0995e-2	21.7282	3	14	1.39%
TS22_optic nerve	7	3.2755e-5	4.3901e-2	2.8067	20	12.74%	28	1.4625e-3	3.8813	16	418	7.41%
TS22_optic stalk	8	4.2295e-5	4.9602e-2	2.7554	20	12.74%	30	1.5399e-3	3.8445	16	422	7.41%
TS17_1st branchial arch	9	4.4572e-5	4.6464e-2	2.3932	25	15.92%	9	3.2638e-4	4.0221	19	479	8.80%

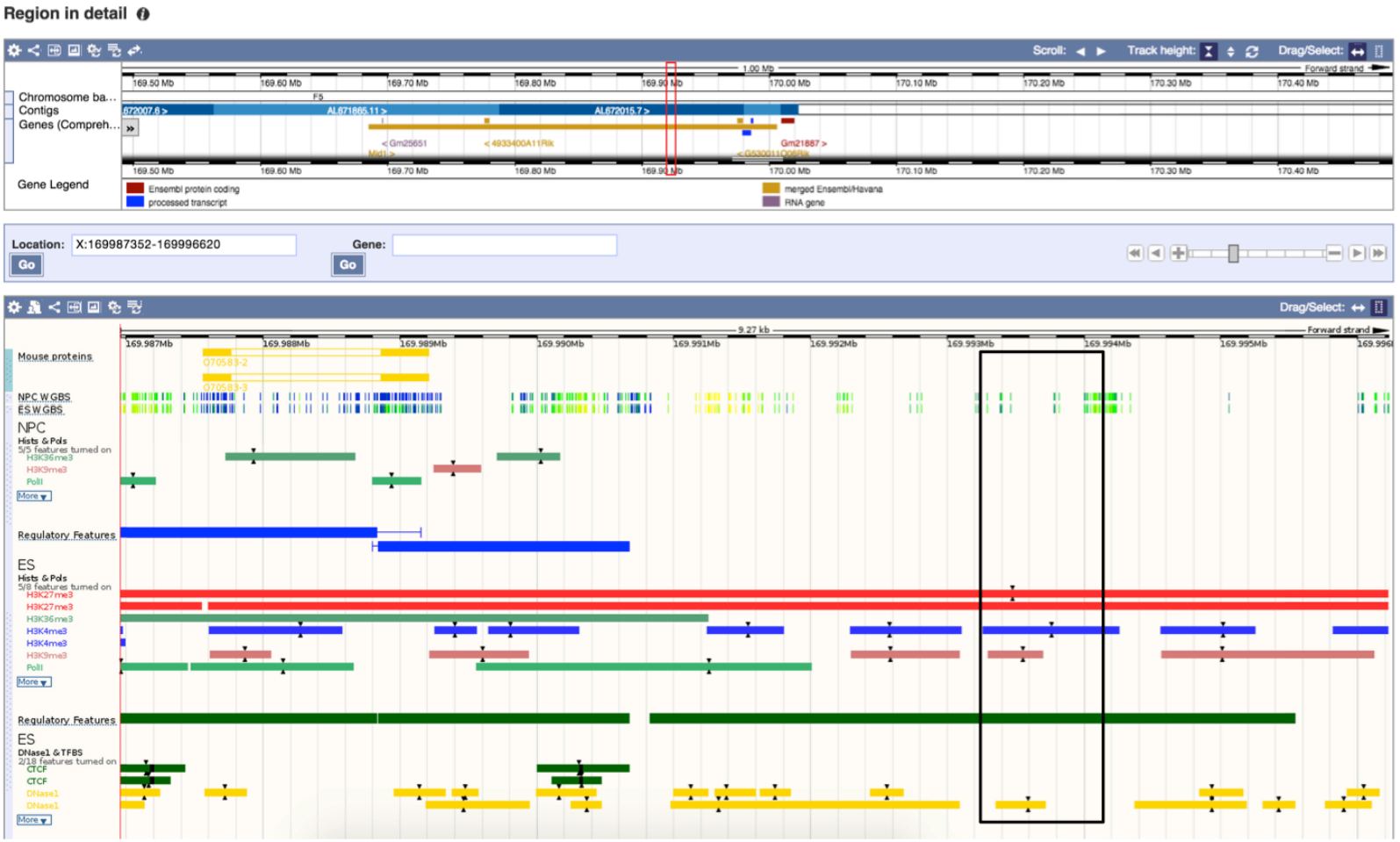
Supplementary Figure 11: Highest-ranked tissue expression terms according to binomial p-value for hyper-DMRs identified between mNSCs and astrocytes using GREAT (v3.0.0). Enrichment for the term "hippocampus granule cell layer" was due to five hyper-DMR regions associated with the genes Tenm2, Tenm4 and Usp29. The two teneurin transmembrane (TENM) proteins are known to be involved in neural development, anterior-posterior axis formation and are most highly expressed in the brain tissue of adult mice (Feng et al., 2002).



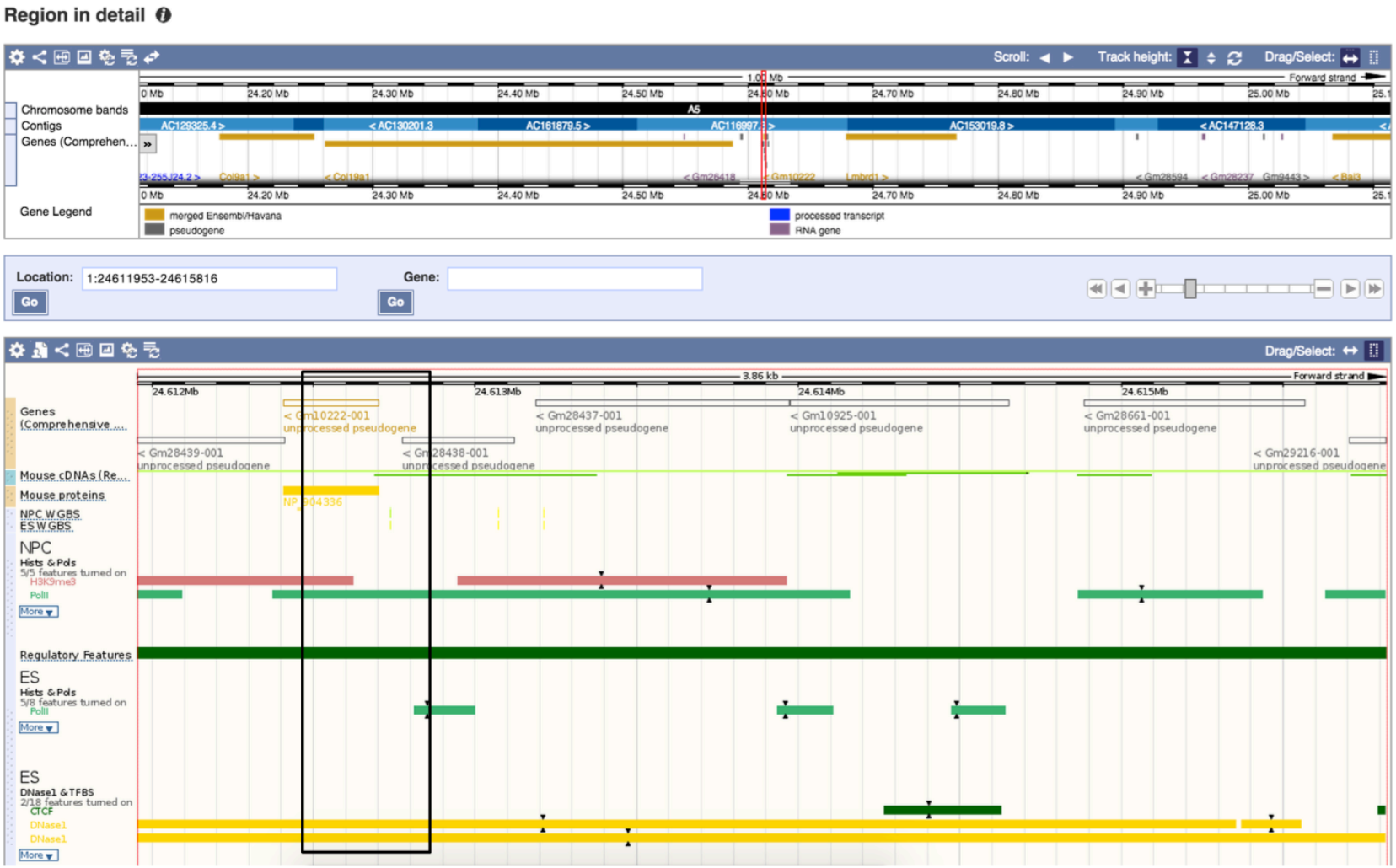
Supplementary Figure 12: Ensembl genome browser view of chr7: 5,020,376-5,033,223. The DNA region identified as the most statistically significant hyper-DMR between mNSCs and mESCs by MeDIP-seq is highlighted by a black box, corresponding to the same DMR highlighted in Figure 4-7. Ensembl regulatory tracks of whole genome bisulphite sequencing (WGBS) data from mESC and mNPC lines confirm this region to be hypermethylated (indicated by dark blue vertical lines) in mNPCs and hypomethylated (indicated by pale green-yellow vertical lines) in mESCs. This DMR region is also associated with both repressive (H3K27me3) and activating (H3K4me3) histone modifications in mESCs only.



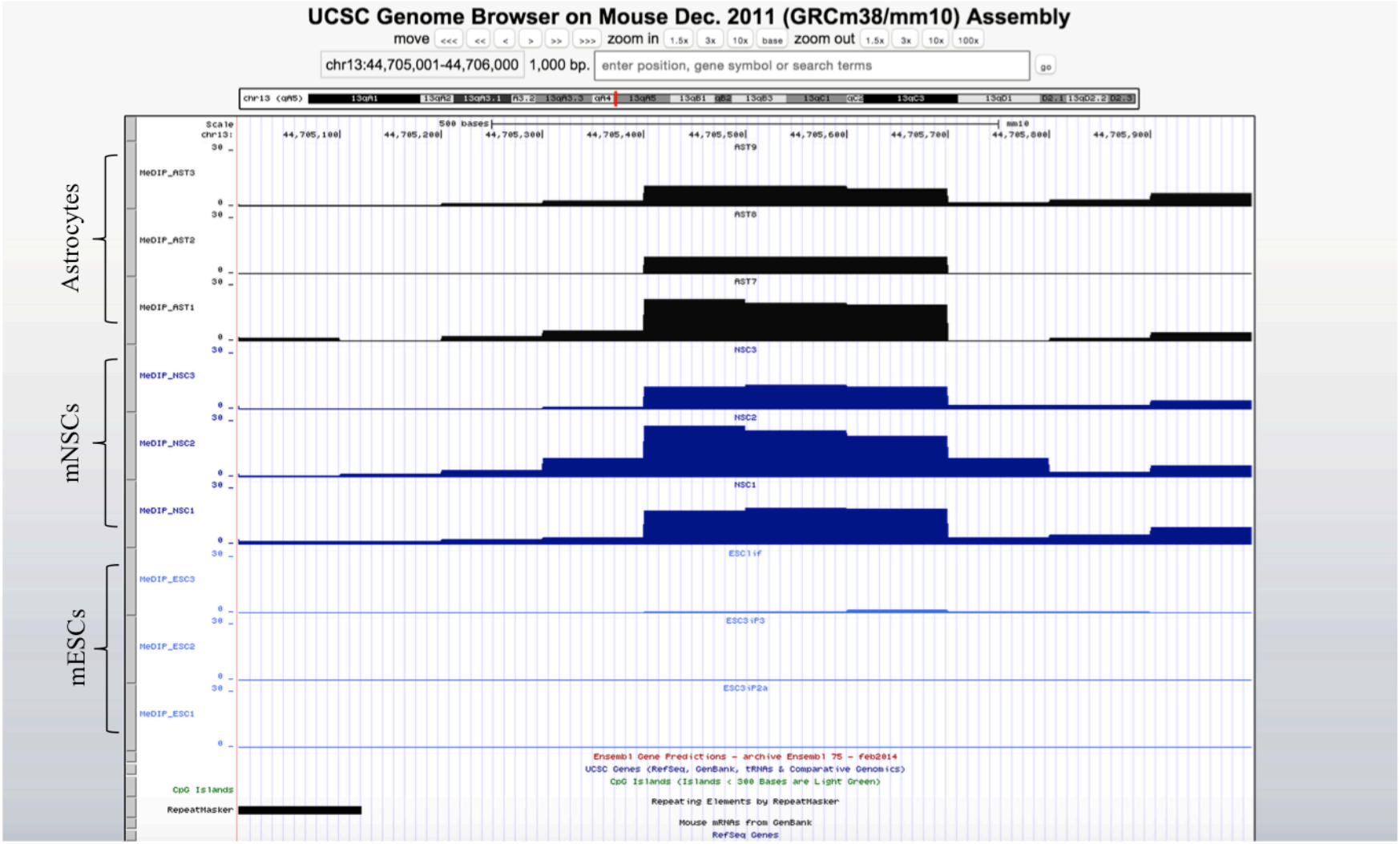
Supplementary Figure 13: Ensembl genome browser view of chr7:49,118,733-49,122,164. The DNA region identified as a hypo-DMR between mNSCs and mESCs by MeDIP-seq is highlighted by a black box, corresponding to the same DMR highlighted in Figure 4-8. Ensembl regulatory tracks of whole genome bisulphite sequencing (WGBS) data from ESC and NPC lines confirm this region to be hypomethylated (indicated by pale green-yellow vertical lines, above) in NPCs compared to ESCs, in which the region is highly methylated (indicated by dark blue vertical lines). In addition, the DMR sits within a region associated with the repressive histone modification H3K27me3 in ESCs only (red bar, above).



Supplementary Figure 14: Ensembl genome browser view of chrX:169,986,956-169,996,224. The DNA region identified as a hyper-hDMR between mNSCs and mESCs by hMeDIP-seq is highlighted by a black box, corresponding to the same hDMR highlighted in Figure 4-9. Ensembl regulatory tracks show that the hDMR sits within a region associated with the repressive histone modifications H3K27me3 (red) and H3K9me3 (pink) as well as the active medication H3K4me3 (blue) in ESCs. The region also sits on a DNaseI hypersensitive site (yellow). WGBS data suggests the region is moderately methylated in both ESCs and NPCs.

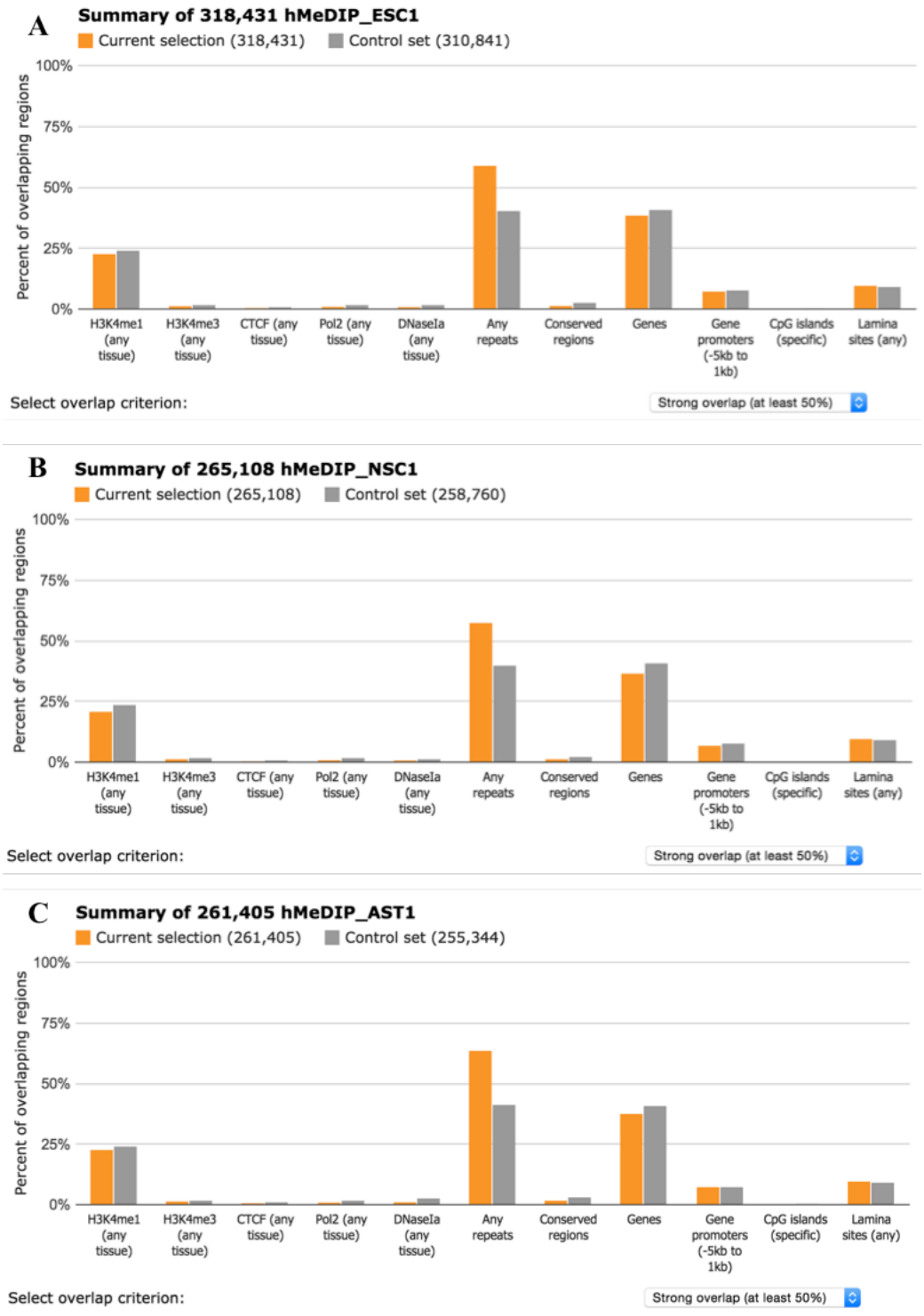


Supplementary Figure 15: Ensembl genome browser view of chr1:24,611,953-24,615,816. The DNA region identified as a hypo-hDMR between mNSCs and mESCs by hMeDIP-seq is highlighted by a black box, corresponding to the same hDMR highlighted in Figure 4-10. Ensembl regulatory tracks show that the hDMR sits within a DNaseI hypersensitive site in mESCs (yellow), whereas in mNPCs the region overlaps a Pol II-enriched site (green) and partially overlaps with H3K9me3 (red).



Supplementary Figure 16: UCSC genome browser view of a hyper-DMR identified between mNSCs and mESCs, approximately 2kb upstream of Jarid2, at chr13:44,705,001-44,706,000. The region appears fully unmethylated in mESCs and partially methylated in both mNSCs and astrocytes.

Supplementary Figure 17: Correlation matrix of beta values for 12 samples interrogated on the Infinium 450K array. The correlation coefficients (r) of all sample pairings are listed above. A correlation coefficient of 1 indicates identical beta value measurements across all 450K probes for the two samples. Only the three 5-aza treatments (0.5, 1.0 and 1.5µM 5-aza) appeared to have any measurable effect on DNA methylation level, indicated by a lower correlation coefficient (0.78-0.79) between 5-aza samples and controls.



Supplementary Figure 18: EpiExplorer analysis of 5hmC peak distribution compared to a randomised control set. A) 5hmC peak distribution in mESCs. B) 5hmC peak distribution in mNSCs. C) 5hmC peak distribution in astrocytes. One representative profile is shown per biological cohort. All hMeDIP profiles were characterised by a greater number of 5hmC peaks within repeat regions than expected by chance, and a very slight depletion of 5hmC peaks within genic region, DNaseI hypersensitivity sites, conserved regions, CTCF and Pol-II binding sites.